UNIVERSIDAD AUTONOMA DE MADRID

Characterization of the regulatory roles of histone H1 in the homeostasis of the mammalian genome

DEPARTMENT OF MOLECULAR BIOLOGY

FACULTY OF SCIENCES

José Miguel Fernández Justel

Madrid, September 2019

UNIVERSIDAD AUTÓNOMA DE MADRID

DEPARTMENT OF MOLECULAR BIOLOGY

FACULTY OF SCIENCES

Characterization of the regulatory roles of histone H1 in the homeostasis of the mammalian genome

José Miguel Fernández Justel

Graduate in Biotechnology

Thesis Director: Dr. María Gómez Vicentefranqueira

Centro de Biología Molecular Severo Ochoa (UAM/CSIC)

Madrid, September 2019

Table of contents

Abbreviations

ADP	Adenosine diphosphate
ARS	Autonomous replicating sequence
ATAC	Assay for transposase accessible chromatin
ATP	Adenosine triphosphate
BSA	Bovine serum albumin
bp	base pairs
CDK	Cyclin-dependent kinase
cheRNA	Chromatin enriched RNA
ChIP	Chromatin immunoprecipitation
CldU	Chlorodeoxyuridine
CPSF	Cleavage and polyadenylation specificity factor
CTD	Carboxy-terminal domain
DDK	Dbf4-dependent kinase
DMEM	Dulbecco's modified Eagle medium
DMR	Differentially methylated region
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
dNTP	Deoxynucleotide triphosphate
DRB	5,6-Dichloro-1-β-D-ribofuranosylbenzimidazole
dsDNA	Double-stranded DNA
dsRNA	Double-stranded RNA
DTT	Dithiothreitol
EDTA	Ethylenediaminetetraacetic acid
EGTA	Ethylene glycol aminoethyl ether tetraacetic acid
eRNA	Enhancer RNA
FACT	Facilitates chromatin transcription
FBS	Fetal bovine serum
FRAP	Fluorescence recovery after photobleaching
H1-TKO	Histone H1 triple knock-out
H3K27ac	Histone H3 lysine 27 acetylation
H3K27me3	Histone H3 lysine 27 trimethylation
H3K4me1	Histone H3 lysine 4 monomethylation
H3K4me3	Histone H3 lysine 4 trimethylation
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
IAS	Internal antisense transcript
IdU	Iododeoxyuridine
KD	Knock-down
LIF	Leukemia inhibitory factor
lncRNA	Long non-coding RNA
mESC	Mouse embryonic stem cells
miRNA	Micro RNA
MNase	Micrococcal nuclease
mRNA	Messenger RNA
ncRNA	Non-coding RNA
NDR	Nucleosome depleted region

NELF	Negative elongation factor
NEXT	Nuclear exosome targeting complex
NHEJ	Non-homologous end joining
ORC	Origin recognition complex
ORI	Replication origin
PAF1C	Polymerase-associated factor 1 complex
PAS	Poly-adenylation site
PAXT	Poly(A)-tail exosome targeting complex
PBS	Phosphate buffered saline
PCR	Polymerase chain reaction
PIC	Pre-initiation complex
PMSF	Phenylmethylsulfonyl fluoride
PNK	Polynucleotide kinase
PRC2	Polycomb repressive complex 2
pre-IC	Pre-initiation complex
pre-RC	Pre-replicative complex
PROMPT	Promoter upstream transcript
PTM	Post-translational modification
ptRNA	Prematurely terminated RNA
P-TEFb	Positive transcription elongation factor b
RBP	RNA binding protein
RNA	Ribonucleic acid
RNApolII	RNA polymerase II
RNP	Ribonucleoprotein
RPKMs	Reads per kilobase per million mapped reads
rRNA	Ribosomal RNA
RT-qPCR	Reverse transcription quantitative PCR
SDS	Sodium dodecyl sulfate
SNS	Short nascent strand
shRNA	Short hairpin RNA
siRNA	Small interfering RNA
snoRNA	Small nucleolar RNA
snRNA	Small nuclear RNA
ssDNA	Single stranded DNA
TAD	Topologically associated domain
TRAMP	Trf4/Air2/Mtr4p polyadenylation complex
tRNA	Transfer RNA
TSS	Transcription start site
TTS	Transcription termination site
uasRNA	Upstream antisense RNA
WT	Wild type

Abstract

Histone H1 is a key component of chromatin, involved in the formation of condensed structures that are refractory to the binding of different factors, thus hindering the action of both the transcription and replication machineries. Previous studies in our group have demonstrated that lowering the total amounts of histone H1 in mouse embryonic stem cells has drastic consequences on their DNA replication dynamics, due to a lack of coordination with transcriptional processes. Along this work, we have continued this investigation, characterizing the mechanisms that account for the conflicts between transcription and replication in H1 deficient cells.

We have found that the replication dynamics alterations are coupled to a perturbed DNA replication initiation landscape, which would also be compatible with massive replication fork stalling mediated by conflicts with the transcriptional machinery. Indeed, we have underscored that the lack of histone H1 produces numerous transcriptional alterations. The most remarkable of them would be the widespread accumulation of unstable non-coding transcripts in chromatin, including PROMPTs, lncRNAs and enhancer RNAs. These transcripts are not post-transcriptionally bound to DNA: their attachment to chromatin is mediated by RNA polymerase II molecules. Coding transcription is also affected, since protein-coding genes display elongation failures and transcriptional read-through. However, all these phenotypes do not seem to be directly related to a chromatin conformational change or any epigenetic alteration, but rather to a defect in RNA processing and metabolism. Finally, we have found that the general phenotype of H1 deficient cells resembles the mutation of several components of the nuclear exosome complex, suggesting a possible functional connection between the linker histone with RNA surveillance mechanisms. These findings have broad implications for our knowledge about the interplay between chromatin and basic cellular processes, and how histone H1 contributes to the maintenance of genome stability and cell homeostasis.

Resumen

La histona H1 es un componente esencial de la cromatina necesaria para la compactación de su estructura, haciéndola así refractaria a la unión de diferentes factores y obstaculizando la acción de las maquinarias de replicación y de transcripción. Estudios previos de nuestro grupo han demostrado que reducir la cantidad global de histona H1 en células embrionarias de ratón tiene consecuencias drásticas en la dinámica de replicación del DNA, debido a la falta de coordinación con procesos transcripcionales. A lo largo de este trabajo, hemos continuado esta investigación, caracterizando los mecanismos que explican los conflictos entre la replicación y la transcripción en células con cantidades reducidas de histona H1.

Hemos descubierto que las alteraciones en la dinámica de replicación están asociadas a un profundo cambio en el paisaje de iniciación de la replicación del DNA. Este fenotipo sería compatible con un bloqueo masivo de las horquillas de replicación, causado por conflictos con la maquinaria transcripcional. En efecto, también hemos comprobado que la falta de histona H1 produce numerosas alteraciones transcripcionales. La más llamativa sería la acumulación en cromatina de transcritos inestables no codificantes, entre los que se encuentran PROMPTs, lncRNAs y "enhancer" RNAs. Estos transcritos no se unen a la cromatina post-transcripcionalmente, sino a través de la RNA polimerasa II. La transcripción codificante también está afectada: los genes codificantes presentan fallos en la elongación y "read-through" transcripcional. Sin embargo, todos estos fenotipos no parecen guardar una relación directa con un cambio conformacional de la cromatina ni con ninguna alteración epigenética, sino más bien con un defecto del procesamiento y el metabolismo del RNA. Finalmente, hemos descubierto que el fenotipo general de las células defectivas para H1 se asemeja al de mutantes en varios componentes del complejo exosoma, lo que sugiere una posible conexión funcional entre la histona H1 y los mecanismos de vigilancia del RNA. Estos hallazgos tienen importantes implicaciones para nuestro conocimiento de la relación entre la cromatina y procesos celulares básicos, y de cómo la histona H1 contribuye al mantenimiento de la estabilidad genómica y la homeostasis celular.

INTRODUCTION

1. DNA replication

DNA duplication is a central biological process, whose fidelity is essential to ensure the correct inheritance of the genome during cell division. To enable the accession of the replication machinery and the onset of DNA synthesis, the DNA double helix structure has to be unwinded. The regions of the genome where this opening takes place are called replication origins (ORIs). In eukaryotes, replication starts from tens of thousands of these replication origins in each cell cycle.

Before replication starts, ORIs undergo two sequential processes: licensing, which consists in the recognition of the origin by the pre-replicative complex (pre-RC), and firing, which leads to the activation of the DNA synthesis. These two processes are separated during the cell cycle: licensing takes place in G1 phase, while firing happens in S phase. This temporal separation is essential to prevent the re-replication within the same cell cycle.

The first step during origin licensing is the recruitment of the origin recognition complex (ORC), a heterohexameric complex of six subunits (from ORC1 to ORC6) with ATPase activity, which is well conserved among all eukaryotes (Bell, 2002; Bell and Dutta, 2002). The genetic features that determine the recognition of an origin by the ORC complex, and hence its replication initiation activity, are still not fully understood. In *Saccharomyces cerevisiae*, ORIs contain an AT-rich 11 bp consensus sequence which, together with other proximal elements, constitutes the ARS (autonomous replicating sequence). ORC complex directly recognize this consensus sequence (Rao and Stillman, 1995), but there are also epigenetic mechanisms that affect the efficiency of the ORC binding (Eaton *et al.*, 2010; Hoggard *et al.*, 2013; Knott *et al.*, 2012). In metazoa, replication origins do not display any consensus sequence; however, several sequence features have been proposed to play a role in their activity, including regions with strand asymmetry (Touchon *et al.*, 2005), CpG islands (Delgado *et al.*, 1998, Sequeira-Mendes *et al.*, 2009), G-quadruplexes (Besnard *et al.*, 2012; Valton *et al.*, 2014), transcription start sites (Cadoret *et al.*, 2008; Cayrou *et al.*, 2011; Dellino *et al.*, 2013) and G-rich elements (Cayrou *et al.*, 2015; Prorok *et al.*, 2019). In addition to these sequence determinants, there are epigenetic factors that affect origin activity, including the presence of chromatin modifiers (Feng *et al.* 2016; Miotto and Struhl 2010; Tardat *et al.*, 2010; Hassan-Zadeh *et al.*, 2012), histone post-translational modifications (Kuo *et al.*, 2012) and nucleosome configuration (Lubelsky *et al.*, 2011; Lombraña *et al.*, 2013).

After ORC assembly into chromatin, the licensing of the ORIs is completed with the association of other proteins: CDC6, CDT1 and, finally, a dimer of two hexamers of the replicative helicases MCM2-7, disposed in an inactive head-to-head conformation encircling dsDNA (Remus *et al.*, 2009). The resulting complex is known as the pre-RC. The assembly of this complex takes place during the mitotic exit and the G1 phase, since it requires the low cyclin-dependent kinases (CDK) conditions that are present at these stages of the cell cycle (Diffley, 2004). During G1-S transition, the increase in CDK and Dbf4-dependent kinase (DDK) activities has two effects on replication regulation: it inhibits the formation of new pre-RC complexes (Petersen *et al.*, 1999; Sugimoto *et al.*, 2004; Chen and Bell, 2011) and initiates the firing of the already licensed ORIs by inducing the phosphorylation of several pre-RC components (Sheu and Stillman, 2006; Francis *et al.*, 2009; Heller *et al.*, 2011), thereby contributing to the temporal separation between both processes. Origin firing requires the assembly of the pre-initiation complex (pre-IC), formed by the MCM complex, GINS and CDC45 (Zou and Stillman, 1998). After the recruitment of other factors, the polymerases are loaded to the complex, the MCM2-7 helicases are activated and bidirectional DNA synthesis starts.

The number of origins that fire during one cell cycle in each cell is much lower than the total number of licensed origins (DePamphilis, 1993). The choice of the origins that are finally activated is variable among the cells in the population: this flexibility could be important to allow the cell to adapt to different environments or cell fate commitments (Mechali, 2010). In most somatic cells, only 10%-20% of the potential origins actually initiate DNA replication (Cayrou *et al.*, 2015), while the remaining ORIs remain dormant and are passively replicated from adjacent replication forks. The presence of these dormant origins is essential for the preservation of genome integrity (Blow *et al.* 2011, Kawabata *et al.* 2011): they serve as backup origins in case that active replication forks stall and are not sufficient to complete replication. There are many different sources of stress that lead to replication forks slowing or stalling, including genotoxic agents, repetitive sequences, G-quadruplex structures, telomeres, DNA-RNA hybrids, errors in the incorporation of dNTPs during replication, transcription-replication conflicts, deregulation of origin activity or reductions of the dNTP pool (Courtot *et al.*, 2018). In all these cases, the cell triggers several mechanisms to repair the DNA damage and complete the genome duplication by firing extra dormant origins, mainly orchestrated by the activation of the ATM-Chk2 and ATR-Chk1 pathways (Ge *et al.*, 2007; Ibarra *et al.*, 2008; Ge and Blow, 2010). Beyond

dormant origin firing, the replicative stress response also involves the protection of the stalled replication fork to avoid its permanent collapse (Lopes *et al.*, 2001; Cobb *et al.*, 2003; Ragland *et al.*, 2013), so DNA synthesis can be subsequently reinitiated through re-priming events (Lopes *et al.*, 2006; Elvers *et al.*, 2011; Mourón *et al.*, 2013), by the recruitment of translesion polymerases (Yeeles *et al.*, 2013), or by template switching mechanisms (Branzei, 2011).

1.1. Transcription-replication conflicts

Replication and transcription are two vital processes for the cell during which the DNA template is copied into complementary DNA or RNA molecules. If both processes take place at the same time, they will potentially compete by the same DNA template. Several studies have demonstrated that the replication forks are stopped when they encounter ongoing transcription (French, 1992; Mirkin and Mirkin, 2005; Prado and Aguilera, 2005), and that regions that are highly transcribed impede fork progression (Azvolinsky *et al.*, 2009). This stalling of the replication fork can lead to the fork collapse and double strand breaks, constituting a major source of replicative stress and genomic instability.

The cells have evolved different strategies to avoid these conflicts between the replication and the transcription machinery. One of them is the physical separation of both processes. In mammalian cells, nucleoli replication and transcription takes place in spatially separated domains (Smirnov *et al.*, 2014), and certain mRNAs that need to be transcribed during early replication are replicated late in S phase, and vice versa (Meryet-Figuiere *et al.*, 2014). However, this spatio-temporal separation is not possible for all genes, including very large ones whose transcription cycle is longer than one cell cycle (Tennyson *et al.*, 1995). Another strategy to minimize the conflicts is the co-orientation of replication and transcription, since collisions in head-on orientation are more harmful than co-directional encounters (Prado and Aguilera, 2005). This co-orientation is clearly observed in bacterial genomes (Merrikh *et al.*, 2012), although it is not so obvious in eukaryotic genomes. The preferential location of replication origins around the transcription start site (TSS) of active genes in mammalian cells (Sequeira-Mendes *et al.*, 2009; Cadoret *et al.*, 2008; Dellino *et al.*, 2013; Chen *et al.*, 2019) could have a role in avoiding head-on collisions, as suggested by the presence of large unidirectional replication zones enriched in highly expressed co-oriented genes (Petryk *et al.*, 2016).

Transcriptional activity can induce the replication fork stalling in several ways. The most direct one is mediated by the presence of the RNA polymerase II (RNApolII) itself, which embraces dsDNA (Barnes *et al.*, 2015), preventing the advance of the replicative helicases. When a molecule of RNA polymerase II encounters a replication fork it can be removed from the template: it has been demonstrated that the retention of RNApolII in yeast chromatin can trigger replication defects (Felipe-Abrio *et al.*, 2015). Apart from this direct effect of RNApolII as a roadblock for the progression of the fork, transcription can impair DNA replication through other indirect mechanisms. During elongation, positive supercoiling is generated in front of the transcription machinery, and negative supercoiling accumulates behind (Liu and Wang, 1987). While the positive supercoiling could directly pose an obstacle to the advance of the replication fork (Tuduri *et al.*, 2009), negative supercoiling facilitates the formation of R-loops (Manzo *et al.*, 2018).

R-loops are nucleic acid structures formed when a RNA molecule invades double stranded DNA, resulting in a RNA-DNA hybrid. This invasion displaces the non-hybridized DNA strand as single stranded DNA, which is more susceptible to DNA damage and prone to the formation of DNA non-canonical secondary structures, like hairpins (Loomis *et al.*, 2014) and G-quadruplexes (Duquette *et al.*, 2004). R-loops are naturally formed during transcription and play several physiological roles in the cell (Skourti-Stathaki and Proudfoot, 2014); however, they also can have deleterious effects through the induction of replication fork stalling.

2. Transcription

Transcription of DNA to RNA is an essential process for all living organisms. In eukaryotes, this process requires (at least) three multi-subunit RNA polymerases: RNApolI, II and III. These three polymerases are responsible for the transcription of different target genes: RNApolI synthesizes the rRNA precursor, RNApolII is involved in the generation of mRNA and several types of non-coding RNAs (lncRNAs, snRNAs, snoRNAs and miRNAs), and RNApolIII transcribes tRNA and 5S rRNA. Plants have evolved two extra polymerases, RNApolIV and V, specialized versions of RNApolII involved in gene silencing.

A transcription cycle is composed of three phases: initiation, elongation and termination. After these three phases are completed, the RNA polymerase can engage in a new cycle. During transcription initiation in metazoans, RNA polymerase II assembles with the general transcription factors (TFIIB, TFIID, TFIIE, TFIIF and TFIIH) at the gene promoter to form the pre-initiation complex (PIC), which is able to open DNA and initiate RNA synthesis (Sainsbury *et al.*, 2015). The assembly of the PIC does not guarantee productive transcription by itself, since RNApolII can pause and accumulate at very high levels in the promoter proximal region, around 30-60 bp downstream the TSS (Adelman and Lis, 2012; Figure 1). The primary function of this pausing of the polymerase is the regulation of the elongation of certain genes, specifically of those involved in stress response and development (Gaertner and Zeitlinger, 2014), since it allows a rapid transcriptional response to different stimuli. Nevertheless, it is also important for the maintenance of a permissive landscape at the promoter (Gilchrist *et al.*, 2008). The release of the polymerase from this region is mediated by the action of the positive transcription elongation factor b (P-TEFb) complex (Peterlin and Price, 2006), which is regulated by different transcription factors, epigenetic mechanisms and splicing factors (Chen *et al.*, 2018). P-TEFb phosphorylates the carboxy-terminal domain (CTD) of RNApolII (besides of other cofactors), allowing it to engage into productive elongation.

During the entire transcription process, the CTD of the RNApolII interacts with different RNA processing factors, depending on its phosphorylation state (Hsin and Manley, 2012): it recruits the 5' capping enzymes shortly after the initiation step, and is essential for the tethering of splicing factors at the elongation step, ensuring the correct cotranscriptional splicing of the RNA. During transcription termination of mRNAs, the RNApolII is slowed down when passes across the polyadenylation site (PAS) of the gene (Figure 1). This facilitates the interaction between the CTD of the polymerase with the cleavage and polyadenylation specificity factor (CPSF), which cleaves the mRNA and process its 3' end (Proudfoot, 2016). The PAS recognition and CPSF assembly can trigger a conformational change in the RNApolII that causes a spontaneous release (Zhang *et al.*, 2015). Still, there is another mechanism to ensure termination that involves the action of the 5'-3' exonuclease Xrn2: Xrn2 is recruited to the PAS and degrades the downstream transcript until it catches up with RNApolII, disassembling it from the DNA template (Proudfoot, 1989). During all the transcription cycle, the RNA is coated with RNA-binding proteins (RBPs) that enable RNA processing and export (Hocine *et al.*, 2010). In most mRNAs, the polyadenylation performed by the CPSF

stimulates the binding of RBPs, generating ribonucleoprotein particles (RNPs) that are competent for nuclear export. Conversely, RNAs that are marked for degradation, instead of export, are not processed by the action of CPSF. Among them are pre-mRNAs with splicing defects, and several types of non-coding transcripts. In yeast, they are oligoadenylated by the Trf4/Air2/Mtr4p polyadenylation (TRAMP) complex, what targets them for nuclear retention and degradation, instead of cytoplasmic export (Tudek *et al.*, 2018); in mammals, the coupling of transcription and degradation of these RNAs is performed by the poly(A)-tail exosome targeting (PAXT) and the nuclear exosome targeting (NEXT) complexes (Meola *et al.*, 2016, Lubas *et al.*, 2015).

To control the final levels of the RNA product, transcription can be regulated in many different ways. The best characterized regulation of transcription takes place at the initiation stage. It involves the action of specific transcription factors, which recognize a consensus sequence or are recruited by other proteins or epigenetic features. Once they are located at the promoter region, they can recruit the transcription machinery or change the chromatin environment around the TSS of target genes. Besides transcription factors, several chromatin features have been shown to regulate the transcriptional activity of a promoter, like the nucleosomal configuration (Jiang and Pugh, 2009), the presence of labile histone variants (Jin *et al.*, 2009) and different histone post-translational modifications (Dong and Weng, 2013).

In addition to their role at initiation, transcription factors and epigenetic features also regulate other stages of the transcription cycle. RNApolII promoter pausing can be regulated by the positioning and stability of the +1 nucleosome (Jimeno-González *et al.*, 2015a; Weber *et al.*, 2014), and P-TEFb (the factor that triggers the release) is recruited and activated by several transcription factors at specific promoters (Rahl *et al.*, 2010; Barboric *et al.*, 2001). The transcription rate during elongation is also regulated: there are several factors that can alter the velocity of the RNA polymerase II, including histone marks that tighten or loosen DNA binding to nucleosomes, histone chaperones or elongation factors (Venkatesh and Workman, 2015; Figure 1). This control of the velocity of the RNApolII determines the cotranscriptional alternative splicing: a decrease on the transcriptional rate can up or down-regulate exon inclusion by the recruitment of different splicing factors (Fong *et al.*, 2014; Dujardin *et al.*, 2014). Conversely, increasing RNApolII elongation rates by altering genomic nucleosomal content causes alternative exon skipping and intron retention (Jimeno-Gonzalez *et al.*, 2015b). Transcription termination and RNA degradation are also used to control gene expression. Premature termination or transcript degradation linked to termination

constitutes a mechanism to prevent or limit gene expression. This is especially important in the case of many non-coding transcripts (Schulz *et al.*, 2013; Schlakow *et al.*, 2017), although the usage of cryptic intronic polyadenylation sites has also been found to regulate the expression of several coding genes (Luo *et al.*, 2013; Kamieniarz-Gdula *et al.*, 2019; Wang *et al.*, 2019).

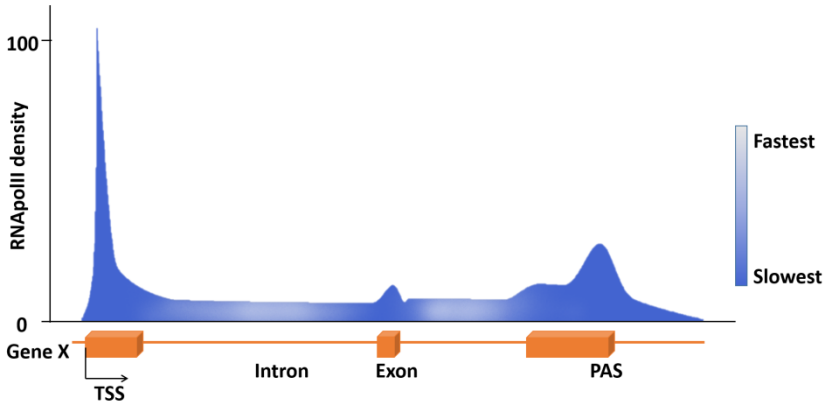


Figure 1: Diagram of the RNA polymerase II occupancy along the body of a gene, obtained from a chromatin immunoprecipitation (ChIP) experiment using an antibody which recognizes total RNApolII (adapted from Jonkers and Lis, 2015). Slow-down or pausing of the polymerase translates into an accumulation of RNApolII and an increase of the ChIP signal at that region. Thereby, there is a clear peak around the promoter region, corresponding to the promoter proximal pausing. Slight accumulations of polymerase can also be detected near exons, where the elongation rate decreases to allow splicing. At the 3' end of the gene, RNApolII signal increases again, when the polymerase is slowed down due the recognition of the poly-adenylation site to enable transcription termination.

2.1. Enhancers

Enhancers are distally located cis-regulatory elements which integrate multiple spatiotemporal stimuli to coordinate gene expression. They act in a tissue-specific manner, and are essential during metazoan development (Levine, 2010). Enhancers were initially described as short DNA fragments that increased the expression of a target gene, independently of their distance or orientation to the promoter. This long-range regulation is thought to occur by a DNA looping mechanism, which brings in close proximity the enhancer with its target promoter (Bulger and Groudine, 2011).

In recent years, large-scale transcriptomic profiling and RNA polymerase II ChIP-seq have demonstrated that enhancers are transcribed, giving rise to a class of non-polyadenylated ncRNAs called enhancer RNAs or eRNAs (De Santa *et al.*, 2010; Kim *et al.*, 2010; Arner *et al.*, 2015). Since then, several studies have demonstrated that eRNAs mark the most functionally active enhancers (Zhu *et al.*, 2013; Hah *et al.*, 2015). In reporter assays, putative enhancers where eRNA transcription had been

described were more likely to trigger reporter activity than those that did not show eRNA transcription (Andersson *et al.*, 2014). Moreover, eRNA-transcribing enhancers display higher binding of transcriptional activators (Kim *et al.*, 2010), greater probability of enhancer-promoter loop formation (Sanyal *et al.*, 2012), increased chromatin accessibility and enrichments in active histone marks like H3K27ac (Hah *et al.*, 2013; Melgar *et al.*, 2011).

In spite of the correlation between eRNA transcription and enhancer activity, it remains controversial whether these RNAs have a role in transcriptional regulation or they are just a byproduct of RNA polymerase II function. On one side, the physical proximity between enhancers and promoters, together with the high chromatin accessibility of enhancers, could trigger an unspecific assembly of the RNApolII machinery, giving rise to an unstable and rapidly degraded RNA. However, several studies support a role of the transcription of eRNAs in enhancer activity. It has been demonstrated that disrupting the transcription of the eRNA in a set of enhancers by introducing early termination signals (Ho *et al.*, 2006; Ling *et al.*, 2004) or by using sh/siRNAs (Melo *et al.*, 2013; Li *et al.*, 2013) and locked nucleic acids (Ilott *et al.*, 2014; Blinka *et al.*, 2016) has a inhibitory effect on the transcription of the coding target gene.

The precise mechanism by which enhancers activate the transcription of their target gene is still under debate. The first model that was proposed was based on the recruitment of several transcription factors to the enhancers: the looping would bring these factors near the promoter, creating the appropriate environment for RNA polymerase II recruitment. Another different mechanism to facilitate transcription initiation depended on the transcription of some eRNAs, which can increase the accessibility of the chromatin of target promoters and stimulate RNApolII binding (Maruyama *et al.*, 2014; Mousavi *et al.*, 2013). However, there is evidence that RNApolII can be stably bound to the promoter before the activation of the enhancer (Zeitlinger *et al.*, 2007; Ghavi-Helm *et al.*, 2014), suggesting that enhancers do not regulate the initiation, but rather the transcriptional elongation. Accordingly, several eRNAs have been shown to play a role in the regulation of the negative elongation factor (NELF, Schaukowitch *et al.*, 2014) and the polymerase-associated factor 1 complex (PAF1C, Chen *et al.*, 2017), two complexes that regulate P-TEFb-dependent RNApolII pause release from the promoter.

2.2. Pervasive transcription

In the human genome, approximately 21000 protein-coding genes have been discovered; however, these protein-coding regions make up less than 3% of the genome length. For many years, the intergenic regions located between these coding genes were thought to be transcriptionally silent "junk DNA", but the advent of high-throughput technologies allowed the identification of many non-coding transcripts that arise from these regions (Bertone *et al.*, 2004; Birney *et al.*, 2007; Mercer *et al.*, 2011; Djebali *et al.*, 2012). Recent estimates indicates that >85% of the human genome is transcribed (Hangauer *et al.*, 2013). Transcripts coming from coding genes only account for 25% of the genomic output, but are one order of magnitude more abundant than ncRNAs.

The non-coding transcripts that are generated in the intergenic part of the genome rise from different regulatory regions. The first of them to be characterized were a set of transcripts that were produced from bidirectionally transcribed CpG-rich promoters (Seila *et al.*, 2008; Ntini *et al.*, 2013). These transcripts were demonstrated to be highly unstable, and were called promoter upstream transcripts (PROMPTs) or upstream antisense RNAs (uasRNAs). Another type of ncRNAs was generated from gene-distal DNaseI-hypersensitive sites (Jacquier, 2009); among them were short transcripts, long non-coding RNAs (lncRNAs) and enhancer RNAs. They are a heterogeneous group of transcripts with different stabilities and processing mechanisms, whose functions are still unknown in many cases. Finally, some of the ncRNAs were originated inside or overlapping the body of coding genes; these RNAs tend to be polyadenylated and accumulated in the nucleus, and regulate the expression of their coding gene partner (Latgé *et al.*, 2018).

Non-coding transcripts can perform different functions in the cell. Besides the enhancer RNAs (whose possible roles have been already described in the previous section), several lncRNAs have been shown to be involved in gene regulation processes, being able to act both in cis and in trans. One of the mechanisms by which they can regulate transcription is though binding and modulating the action of other RNAs or proteins. This is the case for the PANDA ncRNA, which binds the transcription factor NF-YA and prevents it from activating its promoter targets (Kotake *et al.*, 2017), and the CONCR ncRNA, which binds the protein DDX11 and enhances its enzymatic activity during DNA replication (Marchese *et al.*, 2016). lncRNAs can also transport

certain transcription factors or chromatin remodelers to specific targets in the genome. This is the mechanism of action of HOTAIR, which is needed to guide the polycomb repressive complex 2 (PRC2) to the promoter of certain genes and silence them (Yu and Li, 2015). Another described function for lncRNAs is acting like scaffolds, allowing the assembly of multiprotein complexes. One example of this is LINP1, which brings together the proteins Ku80 and DNA-PKcs, so they can act during the non-homologous end joining (NHEJ) repair pathway (Zhang *et al.*, 2016).

Some ncRNAs do not have a function by themselves; the regulatory function is performed by the act of transcription and the presence of elongating RNA polymerase in a specific genomic region (Ard *et al.*, 2017). For example, nearby non-coding transcription can induce the remodeling of the chromatin state of a promoter without the necessity of the ncRNA molecule, thereby increasing its transcriptional activity (Schmitt *et al.*, 2005; Takemata *et al.*, 2016). In other cases, antisense non-coding transcription reduces the level of a mRNA, by transcriptional interference mechanisms (Martianov *et al.*, 2007; Lin *et al.*, 2018).

Pervasive transcription has to be controlled in the cell, degrading many of the useless non-coding transcripts. As mentioned before, in some cases the function is dependent on the act of transcription, not on the transcript per se. Moreover, a pathological accumulation of all these transcripts could be harmful for the cell, risking genome stability or impairing cellular processes. To avoid this, lncRNA transcription and processing is less efficient than that of coding genes due to certain differences during the transcription cycle. In the first place, RNApolII shows a different CTD phosphorylation pattern when transcribing a ncRNA (Schlakow *et al.*, 2017), and it recruits lower levels of elongation factors and higher of termination factors (Fischl *et al.*, 2017; Battaglia *et al.*, 2017). Moreover, the bodies of non-coding genes are enriched in chromatin marks that could difficult transcription elongation (Sun *et al.*, 2015; Murray *et al.*, 2015). However, the main mechanism to decrease the steady-state levels of these non-coding transcripts consists in directly coupling their transcription with the degradation by the nuclear RNA exosome (Mukherjee *et al.*, 2017).

2.3. RNA surveillance by the nuclear exosome

The nuclear exosome is a multiprotein complex that regulates the processing and degradation of different RNA species, transcribed by RNA polymerase I, II and III. It is involved in the maturation and 3' processing of rRNAs, tRNAs, telomeric RNA, snRNAs and snoRNAs (Mitchell *et al.*, 1997; Mitchell, 2014; Zinder and Lima, 2017), and the degradation of transcripts with incomplete 3' ends or defects in polyadenylation (Rondón *et al.*, 2009; Colin *et al.*, 2014), or pre-mRNAs with splicing defects (Bousquet-Antonelli *et al.*, 2000; Lemieux *et al.*, 2011). The RNA exosome is formed by a catalytically inactive core of 9 proteins (EXO-9) and two catalytic subunits: EXOSC10 (or RRP6) and EXOSC11 (or DIS3) (Ogami *et al.*, 2018). Both of them are 3'-5' exonucleases, but EXOSC11 can also act as an endonuclease. In addition to these core proteins, the recruitment of the exosome to their RNA targets requires several additional cofactors (Figure 2). In humans, these cofactors are the TRAMP-like complex, located at the nucleolus and involved in rRNA surveillance (Lubas *et al.*, 2011); and the NEXT and PAXT complexes, which are excluded from the nucleolus and target to the exosome enhancer RNAs (Meola *et al.*, 2016; Lubas *et al.*, 2015), PROMPTs/uasRNAs, prematurely terminated RNAs (ptRNAs) (Ogami *et al.*, 2017), snRNAs (Hrossova *et al.*, 2015), snoRNAs (Lubas *et al.*, 2015) and replication-dependent histone RNAs (Andersen *et al.*, 2013).

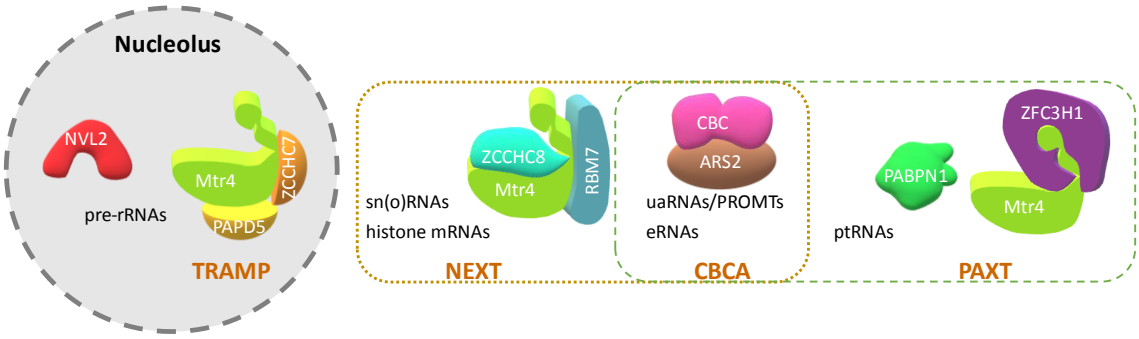


Figure 2: Diagram of the three mammalian exosome cofactors and the RNA categories which are targeted for degradation by each of them (adapted from Ogami *et al.*, 2018).

As stated before, the nuclear exosome RNA degradation activity is essential to regulate the transcription of non-coding transcripts and avoid the harmful effect of their accumulation. Specifically, PROMPTs/uasRNAs and enhancer RNAs are transcripts

with a very low lifetime, since they are degraded by the exosome shortly after their transcription, before they are released from chromatin. This degradation is dependent on the presence of multiple early polyadenylation signals near the TSS, which are recognized by the NEXT complex inducing the premature termination and degradation of the RNA (Ntini *et al.*, 2013; Almada *et al.*, 2013; Andersen *et al.*, 2013; Chiu *et al.*, 2018). Accordingly, when exosome components are mutated or silenced and this regulation layer does not work, both types of transcripts are highly accumulated (Pefanis *et al.*, 2014; Chiu *et al.*, 2018). In addition to eRNAs, other types of lncRNAs are also co-transcriptionally targeted to the exosome, but with a decreased efficiency: on average, they show higher stability (Schlakow *et al.*, 2017).

Chromatin structure and functions

In eukaryotic cells, genomic DNA is packaged through the association with different proteins to form the chromatin fiber. Chromatin has several levels of organization to achieve the sufficient level of compaction to fit the entire genome into a few microns nucleus. The basic repeating structural unit that forms the chromatin is the nucleosome, composed by the core particle and the linker DNA (Kornberg, 1974). The core particle is formed by 147 bp of DNA, wrapped around the histone octamer and completing approximately 1.7 turns around it in a left-handed manner (Luger *et al.*, 1997). Histones are proteins with a highly basic nature, which confers them the ability to bind nucleic acids with high affinity. Two copies of histones H2A, H2B, H3 and H4 form the octamer. At the dyad axis of the nucleosome (the site of DNA entry and exit of the nucleosome core particle), another histone can bind, the linker histone H1, which has a role in the organization of nucleosomes in higher degree levels of compaction (Thoma *et al.*, 1979).

Traditionally, there was a clear distinction between the first two levels of compaction. The first of them was the beads-on-a-string model, or 11nm fiber, consisting just on the nucleosomes disposed in a linear way. This state was characteristic of open euchromatic regions, where transcription was taking place. The addition of histone H1 assembled the nucleosomes in the second level of organization, the 30 nm fiber. This rigid and stable structure was present in transcriptionally silent heterochromatin. Chromatin was further folded into higher-order compaction levels by the formation of loops and coiled structures (Felsenfeld and Groudine, 2003). Recent *in*

vivo studies have questioned this view, establishing a more complex model where the interactions are transient and constantly modified. In particular, nucleosomes have been seen to cluster together in groups of oligonucleosomes (Ricci *et al.*, 2015), with no evidence of regular and stable 30 nm fibers in most cells (Fussner *et al.*, 2012; Chen *et al.*, 2016). In this context, the dynamic association of histone H1 could modulate transient interactions, promoting a constant transition between a condensed and a more opened configuration (Flanagan and Brown, 2016). Accordingly, FRAP experiments revealed that H1 is the most mobile histone, with mean residence times of 3-4 minutes (Lever *et al.*, 2000; Misteli *et al.*, 2000).

Chromatin is not randomly disposed inside the nucleus. At the megabase scale, it is organized into regions with preferential internal interactions called topologically associated domains (TADs). TADs are formed though the action of several proteins, like cohesin and CTCF (Merkenschlager and Nora, 2016), and are involved in the regulation of promoter-promoter and enhancer-promoter interactions (Friedman and Rando, 2015). They associate with other TADs with similar properties to form large compartments of euchromatin and heterochromatin (Schalch, 2017).

Chromatin has essential regulatory functions in all nuclear processes, including transcription and replication (see sections 1 and 2 of the introduction). In general, the chromatin condensation level regulates the accessibility of several factors to the DNA, which define the cellular transcriptional and replicative programmes of the cell (Knott *et al.*, 2009; Hoggard *et al.*, 2013; Feng *et al.*, 2015; Petty and Pillus, 2013; Smolle *et al.*, 2013; Church and Fleming, 2018). The chromatin state is not static, but constantly modulated by the action of chromatin remodelers and histone modifying complexes, which act in response to different stimuli (DesJarlais and Tummino, 2016).

3.1. Histone post-translational modifications

Post-translational modifications (PTMs) of the histones are one of the mechanisms the cell uses to modify the function and the accessibility of the chromatin. Histone tails can be subjected to different PTMs, including acetylation, methylation, phosphorylation, SUMOylation, ubiquitination, mono-ADP-ribosylation and many more (Zhao and García, 2015). One of the first PTMs that were described is histone acetylation. Although these modifications have a complex role which in most cases is not well understood yet, acetylation is known to diminish the electrostatic affinity

between histones and DNA, thereby contributing to the creation of a permissive chromatin structure (Gräff and Tsai, 2013) and facilitating transcriptional activation (Shahbazian and Grunstein, 2007).

Another widely studied modification is the trimethylation in H3K4, which is also a mark of transcriptionally active promoters: the amount of H3K4me3 located at a promoter correlate with the gene expression level (Gu and Lee, 2013). This mark is associated to transcriptional activation through the recruitment of specific transcription factors (Santos-Rosa *et al.*, 2003; Wysocka *et al.*, 2005; Sims *et al.*, 2008), but also by the direct interaction with the basal transcription machinery (Lauberth *et al.*, 2014). Conversely, the presence of monomethylation of H3K4 has been shown to have a negative effect on the transcription of the gene (Cheng *et al.*, 2014). H3K4me1 is a mark traditionally associated to enhancers (Robertson *et al.* 2008), although it is not clear whether this association is merely a correlation or it plays any role in the enhancer function (Catarino and Stark, 2018).

Other PTM that silences a gene when it is present at its promoter is H3K27me3. H3K27 is trimethylated by the polycomb repressive complex 2 (PRC2), formed by the proteins SUZ12, EED and EZH1/2, which is recruited to the promoter of certain genes by DNA target sequences, other histone modifications or non-coding RNAs. Once PRC2 is recruited, it trimethylates H3K27 and inhibit transcription initiation by different mechanisms (Aranda *et al.*, 2015). H3K27me3 and H3K4me3 coexist at a set of specific promoters mainly involved in development; this is called the “bivalent state” (Bernstein *et al.*, 2006). The genes with this combination of histone PTMs are kept poised for subsequent activation or repression in distinct cell types.

3.2. Histone variants

In addition to post-translational modifications, nucleosome properties can also be modulated by the replacement of the canonical histones by histone variants with a slightly modified aminoacid sequence. In metazoan, histone variants show several differences with canonical histones. In the first place, canonical histone genes are present in multiple copies, forming clusters in specific chromosomal locations. Their mRNA is intronless and, instead of the poly-A tail, has a 3’ stem-loop structure that allows them to be exported and translated exclusively during S phase. This cell cycle dependent expression is related to their assembly into chromatin: they are deposited

mainly during replication, in the DNA that is being synthesized. Conversely, histone variant genes do not form clusters, and are transcribed along all the cell cycle to produce intronic and polyadenilated mRNAs. They are incorporated to DNA by different histone chaperones, also in a cell cycle independent manner (Talbert and Henikoff, 2017).

The replacement of canonical histones by variants alters the properties of nucleosomes, changing their interactions with chromatin remodelers and modifiers, and playing different functions. For example, CENP-A substitutes canonical H3 at centromeres, and it is necessary for correct kinetochore assembly and chromosome segregation (Howman *et al.*, 2000); H2A.X is deposited at double strand break sites and, after becoming phosphorylated in Ser139, is essential to trigger DNA damage response and repair pathways (Podhorecka *et al.*, 2010); H2A.Z and H3.3 are enriched around the TSS of active genes, regulating transcriptional activation (Shi *et al.*, 2017; Giaimo *et al.*, 2019); and MacroH2A stabilizes the binding of the nucleosome to the DNA, leading to transcriptional repression (Changolkar and Pehrson, 2006).

3.3. Nucleosome landscape

The assembly of a nucleosomal particle protects the 147 bp of bound DNA from most interactions with other proteins that bind DNA. Thus, the competition between the histone octamer and several protein complexes, including transcription factors or the replication machinery, is essential to regulate different cellular processes. There is certain sequence specificity that determines the nucleosome binding pattern to DNA, although it seems to vary between different organisms (Ioshikes *et al.*, 1996; Schones *et al.*, 2008; González *et al.*, 2016). However, this sequence-guided assembly can be modified by active regulatory mechanisms of the cell, called chromatin remodeler complexes. There are four subfamilies of chromatin remodelers: ISWI, CHD, SWI/SNF and INO80. They regulate nucleosomal configuration in several ways: nucleosome assembly after replication and transcription; regulation of chromatin access by sliding or evicting nucleosomes, which expose binding sites for transcription factors at promoters or enhancers; and nucleosome editing, allowing the replacement of histones by other variants (Clapier *et al.*, 2017).

The nucleosomal configuration of a given genomic position is defined by two parameters: occupancy and positioning (Figure 3). Nucleosome positioning refers to

the exact location of the start of the 147 DNA basepairs that are protected by the octamer particle: a nucleosome is well positioned if a high percentage of the cells in the population display it in exactly the same genomic coordinates. Occupancy is related to the presence or absence of these particles over specific locations: a genomic region shows high occupancy if it is covered by a nucleosome in most of the cells of the population (Arya *et al.*, 2010).

Maps of the nucleosome landscape at a genome-wide scale can be achieved by sequencing mononucleosomal DNA resistant to the action of Micrococcal Nuclease (MNase), which is able to digest the linker DNA but not the octamer bound DNA. One of the main conclusions of the generation of these maps has been the specific pattern of nucleosome configuration present at eukaryotic promoters and enhancers: a nucleosome depleted region (NDR) flanked by two well positioned nucleosomes. The NDR encompass the transcription start site, so the flanking nucleosomes are called the -1 (upstream the TSS) and the +1 (downstream the TSS). This configuration has been shown to be important for the recruitment of transcription factors and the assembly of the transcriptional machinery (Lee *et al.*, 2007; Buratowski, 2008). Downstream the +1 nucleosome, a strong phasing of 5-10 nucleosomes can be detected; then, the positioning is lost inside the gene body (Arya *et al.*, 2010; Figure 3). At the 3' end of the gene, there is another well positioned nucleosome, followed by the 3' NDR. This particular configuration is important to facilitate RNApolII disengagement and transcription termination.

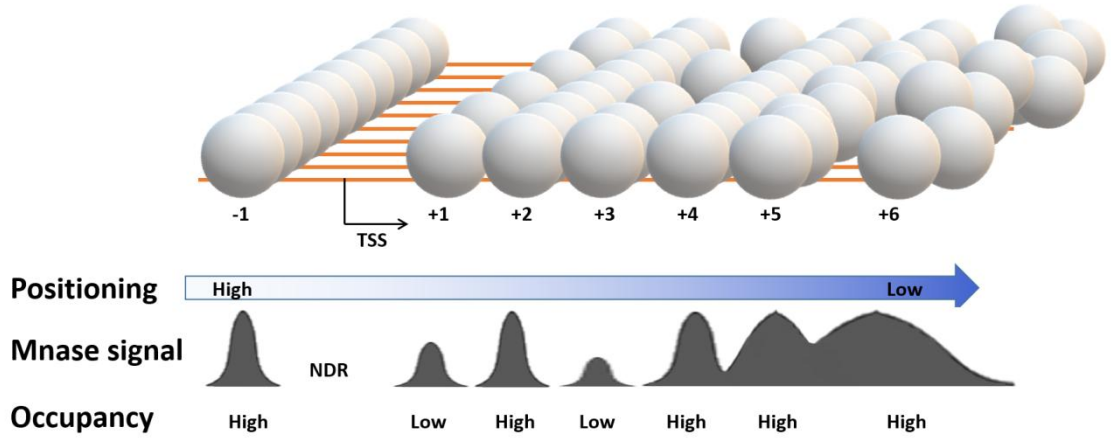


Figure 3: Schematic definition of the nucleosome parameters positioning and occupancy (adapted from Pugh 2010). Each orange line represents the DNA of one single cell within the population, covered by nucleosomes (the grey beads). In the promoter of a gene, there is a nucleosome depleted region matching the TSS. The two flanking nucleosomes are well positioned, but this positioning is gradually lost while entering the gene body.

In addition to transcription, nucleosome configuration is also important to regulate replication initiation. In *S. cerevisiae*, the presence of a NDR within the ARS is required to allow the binding of the ORC complex (Bell *et al.*, 1995); however, this NDR cannot be too wide, so the positioning of the flanking nucleosomes has to play a role in origin specification (Lipford and Bell, 2001). In metazoan, this colocalization between low occupancy regions and replication origins has also been detected (MacAlpine *et al.*, 2010; Lubelsky *et al.*, 2011). In a high-resolution analysis of nucleosome configuration and DNA synthesis initiation in mammalian cells, it was shown that replication initiation at CpG islands promoters matches positions of high occupancy, but the adjacent ORC-binding site coincides with the presence of labile nucleosomes (Lombraña *et al.*, 2013).

4. Histone H1

The linker histone H1 is a key component of chromatin. The binding of H1 to the core nucleosomal particle plays a role in chromatin compaction (see section 3), and reduces the size of the linker DNA: it makes nucleosomes to be closer to each other (Woodcock *et al.*, 2006). However, its precise function has been proven to be the most difficult to understand, and is not completely characterized yet.

Histone H1 family is the less conserved and most heterogeneous group among histone proteins. In mammals, there are at least 11 H1 variants (Table 1), including four that are specific from germ cells (H1t, H1oo, H1T2 and HILS1) and seven that are present in somatic cells. Inside the latter group, there are five replication-independent (H1a, H1b, H1c, H1d and H1e) and two replication-dependent variants ($H1^0$ and H1x). All these variants show distinct location patterns along the genome and perform different cellular roles. Still, they are partially redundant, showing overlapping binding profiles and compensation effects upon single knockouts (Hergeth and Schneider, 2015). They share the core histones features regarding gene organization and mRNA processing, and can also suffer post-translational modifications.

Due to its role in heterochromatin formation in different organisms (Lu *et al.*, 2009; Cao *et al.*, 2013; Popova *et al.*, 2013), histone H1 was thought to reduce the accessibility of the promoters to transcription factors, being a general transcription

repressor. Accordingly, several studies in different models have shown that transcriptional activation of a promoter usually requires the eviction of H1 from it (Breshnick *et al.*, 1992; Braunschweig *et al.*, 2009; Krishnakumar *et al.*, 2008). In fact, transcriptionally active genes show a clear depletion of histone H1 from their promoters, which is also absent from other regulatory regions as enhancers and insulators (Izzo *et al.*, 2013; Millán-Ariño *et al.*, 2014). This eviction is facilitated by several post-translational modifications of the histone induced by distinct transcription factors, including phosphorylation (Zheng *et al.*, 2010), acetylation (Kamieniarz *et al.*, 2012) or poly-ADP-ribosylation (Azad *et al.*, 2018). Certain chromatin modifications, like H3K27me3, prevent H1 displacement, recruiting it to silenced genes (Kim *et al.*, 2015). Despite this requirement of H1 absence at promoters for gene expression, decreasing the overall levels of histone H1 does not trigger a global increase in transcriptional activity. Several H1 knock-out and knock-down models in different organisms showed limited transcriptional alterations, comprising both up and down-regulation of specific sets of genes (Shen and Gorovsky, 1996; Hashimoto *et al.*, 2010; Vujatovic *et al.*, 2012; Fan *et al.*, 2005; Sancho *et al.*, 2008; Izquierdo-Boulstridge *et al.*, 2017).

Protein	Mouse gene	Expression specifity	DNA-replication dependent expression
H1.a (H1.1)	*Hist1h1a*	Somatic	Yes
H1.b (H1.5)	*Hist1h1b*	Somatic	Yes
H1.c (H1.2)	*Hist1h1c*	Somatic	Both*
H1.d (H1.3)	*Hist1h1d*	Somatic	Yes
H1.e (H1.4)	*Hist1h1e*	Somatic	Yes
H1^0 (H1.0, H1f0)	*H1f0*	Somatic (enriched in differentiated cells), oocytes	No
H1.x (H1.x, H1fx)	*H1fx*	Somatic	No
H1oo (H1foo)	*H1foo*	Oocytes, zygote and 2-cell embryo	No
H1t	*Hist1h1t*	Spermatocytes, spermatids	Yes
H1T2	*H1fnt*	Spermatids	No
HILS1	*Hils1*	Spermatids	No

Table 1: Comparative overview of H1 histone variants expression mode and tissue specificity (adapted from Pan and Fan, 2016) *Hist1h1c gene produces two alternative mRNAs, one polyadenylated and one with a 3' stem-loop, allowing the expression of H1.c in dividing and nondividing cells (Cheng *et al.*, 1989).

Histone H1 is involved in other cellular processes, beyond transcriptional regulation. For example, in response to double strand breaks H1 is ubiquitylated, leading to the recruitment of repair factors (Thorslund *et al.*, 2015). During the cell cycle, phosphorylation of H1 by CDK1/CycB could be required for metaphase chromosome compaction (Th'ng *et al.*, 1994; Maresca *et al.*, 2005). Finally, some of the germ line specific variants play different roles during gametogenesis and early embryogenesis (Pan and Fan, 2016).

4.1. H1 triple knock-out mouse embryonic stem cells

To study the function of the linker histone H1 in mammals, several mouse knock-out models of individual variants were generated. However, all of them displayed very slight phenotypes: the mice were viable and fertile, and the cells showed no obvious defects in transcriptional control (Fan *et al.*, 2001; Lin *et al.*, 2000; Sirotkin *et al.*, 1995; Rabini *et al.*, 2000). In these studies, it was described that the lack of any apparent phenotype was likely due to the upregulation of the remaining subtypes, which was enough to maintain a normal linker-to-core stoichiometry. To overcome this compensation mechanism, three histone H1 variants were deleted sequentially in mouse embryonic stem cells (mESCs): H1c, H1d and H1e (Fan *et al.*, 2003).

Mice derived from H1 triple knock-out (H1-TKO) cells were not viable, dying by mid-gestation with a broad range of defects. In posterior studies, H1-TKO mES cells were characterized: they showed a 50% reduction on total histone H1 content, what leads to local reductions in chromatin compaction and a decrease in the nucleosomal spacing (Fan *et al.*, 2005). However, in a similar way than other H1 deficient models, H1-TKO cells do not display global transcriptional alterations: only the expression of a few genes is changed.

In recent studies of our group, we have made use of the H1-TKO cells to investigate how chromatin participates in the coordination between the processes of replication and transcription (Almeida *et al.*, 2018). By analyzing DNA replication dynamics through fiber stretching (see Materials and Methods), we found that the lack of histone H1 causes multiple defects in DNA replication: replication forks move slower (Figure 4a, untreated columns, compare dark blue and dark red plots), what is accompanied by an increase in the number of active replication origins (seen as a

reduction of the inter-origin distance, figure 4b). Moreover, the two forks that rise bidirectionally from a single origin tend to produce asymmetrically replicated DNA tracks (Figure 4c), which is a signal of fork stalling and instability.

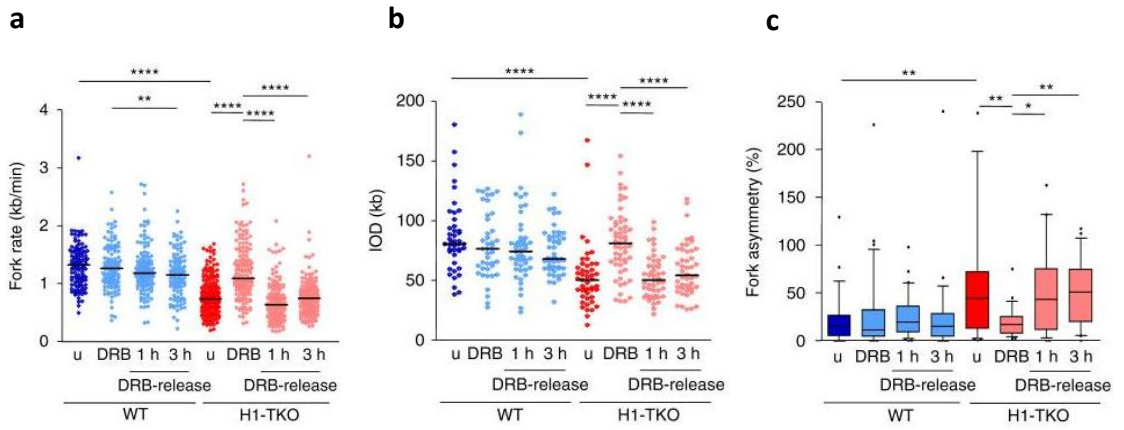


Figure 4: Plots showing the replicative defects in H1-TKO cells. **(a)** Fork rates, **(b)** Inter-origin distances and **(c)** Percentage of fork asymmetry between WT (blue) and H1-TKO (red) cells, untreated (u), treated with DRB for 3h (DRB) and released from the DRB block (1 h and 3 h DRB-release). Median values are indicated (n = 2). Differences between distributions were assessed with the Mann–Whitney rank sum test. ****p-value < 0.0001; ***p-value < 0.001; **p-value< 0.1; *p-value < 0.2. Adapted from Almeida *et al.*, 2018.

Importantly, all these replicative alterations were dependent on the presence of active transcription: the addition of DRB for three hours, a drug that inhibits the promoter pausing release of the RNApolII, completely reverted this phenotype (Figure 4, DRB columns), which is readily re-established upon transcription re-start (Figure 4, DRB release columns). These results suggest that transcription-replication conflicts are the main agent that causes the DNA replication defects in H1-TKO cells.

OBJECTIVES

The general aim of this work is the investigation of the mechanistic link between the linker histone H1 and the coordination of the processes of transcription and replication in H1-TKO mES cells. To understand how the absence of H1 leads to transcription-replication conflicts and replicative stress, we set the following objectives for the Doctoral Thesis:

1. Analyze the specification and activity of DNA replication origins in histone H1 deficient cells.

2. Study the transcriptional alterations triggered by the lack of correct amounts of histone H1 that could explain the conflicts with the replication machinery.

3. Analyze the epigenetic features and the nucleosomal landscape changes upon reduction of H1 levels, and how they are related to the transcriptional and replicative phenotype.

MATERIALS AND METHODS

1. Experimental methods

1.1. Cell culture

Mouse embryonic stem cells were grown in DMEM (Invitrogen) supplemented with 10% fetal bovine serum (Biosera), 1x non-essential aminoacids (Gibco), 1mM sodium piruvate (Gibco), 2mM L-glutamine (Gibco), 50 μM β-mercaptoethanol (Gibco), 10^3 U/mL LIF (ESGRO), 100 U/mL penicillin (Invitrogen) and 100 μg/mL streptomycin (Invitrogen), in a humidified atmosphere at 37ºC and 5% CO_2. Cells were always maintained on a mytomycin C (Sigma) treated mouse embrionary fibroblast feeder monolayer, and a gelatin-coated (Sigma) surface. During transcription inhibition experiments, mES cells at 80% confluency were treated with 100 μM DRB (Sigma) for 3 hours.

293FT cells were grown in DMEM (Invitrogen) supplemented with 10% fetal bovine serum (Gibco), 100 U/mL penicillin (Invitrogen) and 100 μg/mL streptomycin (Invitrogen).

mES and 293T cells were subcultured every 2 days, at a ratio of 1:5-1:7 depending on the confluency. Culture media was aspirated, they were washed twice with PBS and then treated with a solution containing 0.25% trypsin, 0.04% EDTA and 2% chicken serum (Invitrogen) for approximately 5 minutes. Detached cells were collected in fresh culture medium and centrifuged for 5 minutes at 180g. Finally, the cell pellet was resuspended in medium and plated in the new culture surface.

For long-term storage, cells at 80% confluency were trypsinized and collected as described before. Approximately $8 \cdot 10^6$ cells were resuspended in 1 mL of freezing medium, containing 50% mES growth medium, 40% FBS (Biosera) and 10% DMSO (Merck), and aliquoted in a cryovial. They were frozen by placing them in a Mr. Frosty container at -80ºC overnight. The following day the cells were transferred to liquid nitrogen.

During cell reactivation, one cryovial was thawed at 37ºC in a water bath. The cells were diluted in 10 mL of pre-warmed culture medium, and centrifuged at 180g for

5 minutes. The cell pellet was resuspended in fresh medium, and plated in a T25 tissue culture flask.

For shRNA stable integration, 5 p100 plates of 293FT cells at 80% confluency were co-transfected with 24 μg psPAX2 (AddGene plasmid #12260), 7.2 μg pMD2G (AddGene plasmid #12259) and 33 μg pLKO.1 (AddGene empty backbone #8453) plasmids, using Lipofectamine™ 2000 (Thermo) following manufacturer's instructions. At 48h and 72h after transfection, the growing medium was collected and passed through a 0.45 μm pore low binding filter. The lentiviral particles were purified by ultracentrifugation at 85000g and 4ºC during 3 hours, resuspended in PBS, and added to one p24 well of mES cells at 80% confluency. 72h later, the medium was supplemented with 0.5 μg/mL of puromycin. The cells were considered stably transduced after 7 days of antibiotic selection.

1.2. Short nascent strands isolation and sequencing

$100 \cdot 10^6$ exponentially growing mouse embrionary stem cells were lysed by directly adding to the culture plates 5 mL of lysis buffer (50 mM Tris pH 8, 10 mM NaCl, 10 mM EDTA pH 8, 0.5% SDS). The cell lysate was collected, supplemented with 100 μg/mL proteinase K (Roche) and incubated overnight at 37ºC. The following day, it was mixed with an equal volume of Tris-saturated phenol (Sigma) and centrifuged 10 minutes at 1500g. The resulting aqueous phase was further purified by a second extraction with an equal volume of phenol:chloroform:isoamylalcohol (25:24:1). DNA was precipitated with 2 volumes of pre-cooled 100% ethanol (Merck), washed with 70% ethanol, air-dried for 30 minutes and resuspended in 1mL of TE buffer (10 mM Tris pH 8, 1mM EDTA) supplemented with 0.1 U/μL RNAseOUT™(Invitrogen).

The genomic DNA was denatured by heating at 100ºC for 10 minutes and size-fractionated by ultracentrifugation at 78000g for 20 hours at 20ºC in a SW40 rotor (Beckman ultracentrifuge), placing it on top of 12 mL of a seven step sucrose gradient (from 5% to 20% sucrose, in 2.5% steps, diluted in 10 mM Tris pH 8, 1mM EDTA and 100 mM NaCl). After the centrifugation, the gradient was divided in thirteen 1mL aliquots. From each of them, DNA was precipitated with 1:10 volume of 5M NaCl and 2 volumes of pre-cooled 100% ethanol, washed with 70% ethanol, air-dried for 30 minutes and resuspended in 100 μL TE buffer. 10 μL of the volume of these fractions

was analyzed by running it in a 1% alkaline agarose gel (50 mM NaOH, 1mM EDTA) to evaluate the size of the purified DNA. Gradient fractions with DNA fragments showing a range of sizes between 300-1000 bp were selected, and treated with 100 Units of polynucleotide kinase (PNK, Thermo Scientific), in a buffer containing 1mM dATP (Roche) and 40 Units of RNAseOUT™. After 30 minutes of 37ºC incubation, the enzyme was inactivated by adding 6.25 μg of proteinase K, 0.125% sarkosyl and 2.5 μM EDTA, and DNA was ethanol-precipitated. PNK treatment is essential to phosphorylate the free 5' hydroxyl group of the molecules of the sample, making them susceptible to degradation by λ-exonuclease enzyme, which degrades broken DNA leaving RNA-protected leading strands intact. For λ-exonuclease treatment, the DNA pellet was resuspended in water and heat-denaturated for 5 minutes; next, 10x λ-exonuclease digestion buffer, 150 Units of enzyme and 40 Units of RNAseOUT™ were added. After an overnight 37ºC incubation, the reaction was stopped by heating 10 minutes at 75ºC. The resulting DNA was extracted with phenol-chloroform, ethanol-precipitated and resuspended in water. This cycle of PNK + λ-exonuclease treatments was repeated three times.

Before generating sequencing libraries, the short nascent strand intermediates, which were single stranded DNA fragments, were transformed into double stranded DNA, as described in Cadoret *et al.*, 2008. For that, they were primed with 50 pmol of random hexamer primers (Roche), incubated for 5 minutes at 95 ºC and gradually cooled to 4ºC. The primers were extended with a treatment with 5 Units of exo-Klenow (New England Biolabs) and 10 mM dNTPs (Roche), during one hour at 37 ºC. The enzymatic reaction was stopped by heating at 75ºC for 10 minutes. To ligate the resulting fragments, 80 Units of Taq DNA ligase were added, and the samples were incubated for 30 minutes at 50ºC. The enzyme was inactivated by heating at 75ºC during 10 minutes, and RNA primers were removed with a treatment with 5 Units of RNAse A/T1 mix (Thermo Scientific), during 30 minutes at 37ºC. The resulting double stranded DNA was purified by ethanol precipitation.

DNA libraries were prepared with NBNext kit (New England Biolabs) following manufacturer's instructions. Library fragments were purified from polyacrylamide gels and sequenced by 1x75 single-end runs on a Illumina NS500 system, at the Fundación Parque Científico de Madrid.

1.3. Cell fractionation

The cell fractionation was performed as described in Mendez and Stillman, 2000, with some minor modifications. Briefly, $20 \cdot 10^6$ mES cells were trypsinized and resuspended in 500 μL of buffer A (10 mM HEPES pH 7.9, 10 mM KCl, 1.5 mM $MgCl_2$, 0.34M sucrose, 10% glycerol, 1mM DTT, 10 μM leupeptin, 100 μM PMSF, 1μM pepstatin, 2 μg/mL aprotinin, 5 mM NaF, 1mM $NaVO_3$). The cytoplasmic membrane was lysed by the addition of 0.1% Triton X-100, and incubation for 10 minutes on ice. The nuclei were collected by centrifugation for 4 minutes at 1300g, and the cytoplasmic supernatant fraction was stored. The nuclei pellet was washed once in buffer A, and then lysed in 200 μL of buffer B (3 mM EDTA, 0.2 mM EGTA, 1 mM DTT and protease and phosphatase inhibitors as described above). The insoluble chromatin fraction was collected by centrifugation for 4 minutes at 1700g, and washed once with buffer B; the nucleoplasmic supernatant fraction was stored. The chromatin pellet was resuspended in RIPA buffer (25 mM Tris pH 7.5, 180 mM NaCl, 1% NP-40, 1% Na-deoxycholate, 0.1% SDS and protease and phosphatase inhibitors), and sonicated in Bioruptor® for 10 cycles, 30 seconds high and 30 seconds off. Finally, the three fractions were clarified by centrifugation for 15 minutes at 20000g to remove debris and insoluble aggregates.

1.4. Chromatin enriched RNA isolation and sequencing

cheRNA preparations were obtained as described in Werner and Ruthenburg, 2015, with the inclusion of an in vitro transcribed spike-in to allow absolute quantification. $40 \cdot 10^6$ mES cells were trypsinised, counted in a Neubauer chamber and collected by centrifugation. They were washed once with PBS + 1mM EDTA, and their cytoplasmic membrane was lysed by resuspending them in 800 μL ice-cold Lysis Buffer A (10 mM Tris pH 7.5, 0.1% NP-40, 150 mM NaCl). After a 5 minute incubation in ice, the lysate was layered on top of 2.5 volumes of a chilled sucrose cushion (24% sucrose in lysis buffer A), and centrifuged for 10 minutes at 4ºC, 200g. The nuclei pellet was carefully rinsed with ice-cold PBS + 1mM EDTA, and resuspended in 500μL ice-cold Glycerol Buffer (20 mM Tris pH 7.9, 75 mM NaCl, 0.5 mM EDTA, 0.85 mM DTT, 0.125 mM PMSF, 50% glycerol). One volume of ice-cold Lysis Buffer B (10 mM HEPES pH 7.6, 1mM DTT, 7.5 mM $MgCl_2$, 0.2 mM EDTA, 0.3M NaCl, 1M urea, 1% NP-40) was added, and the nuclei were lysed on ice for 10 minutes, with periodic vigorous shaking. The presence of urea in this lysis step is essential, because it

removes weak interactions of RNA with chromatin, leaving almost exclusively RNA polymerase bound transcripts. Insoluble chromatin was sedimented by centrifugation at 15000g and 4ºC for 2 minutes, and the nucleoplasmic supernatant fraction was stored. The chromatin pellet was carefully rinsed twice with cold PBS + 1mM EDTA, and resuspended in 100 μL PBS. At this point, 10 pg of an in vitro transcribed luciferase RNA (kindly provided by Dr. Encarna Martínez-Salas) was added to both the nucleoplasmic and chromatin samples as a spike-in control. The nucleoplasmic and the chromatin associated RNA were purified using TRIzol™, following manufacturer's instructions. Finally, they were subjected to two rounds of treatment with DNaseI (Invitrogen), and purified by phenol-chloroform extraction and ethanol precipitation. The RNA was quantified in Nanodrop™ One, and 500 ng were run in a non-denaturing 1% agarose gel to estimate its quality.

The enrichment in chromatin associated transcripts in cheRNA preparations was monitored by checking the presence of nascent pre-processed ribosomal RNA, which appeared at higher molecular weights than the final 28S and 18S rRNA bands in an agarose gel (Figure 5a). Also, the chromatin/nucleoplasm ratio for two non-coding RNAs which are known to associate to chromatin post-transcriptionally (Kcnq1ot1 and NEAT1) was measured by RT-qPCR, comparing them with two normally exported mRNAs (Klf16 and Nat8L) used as negative controls (Figure 5b).

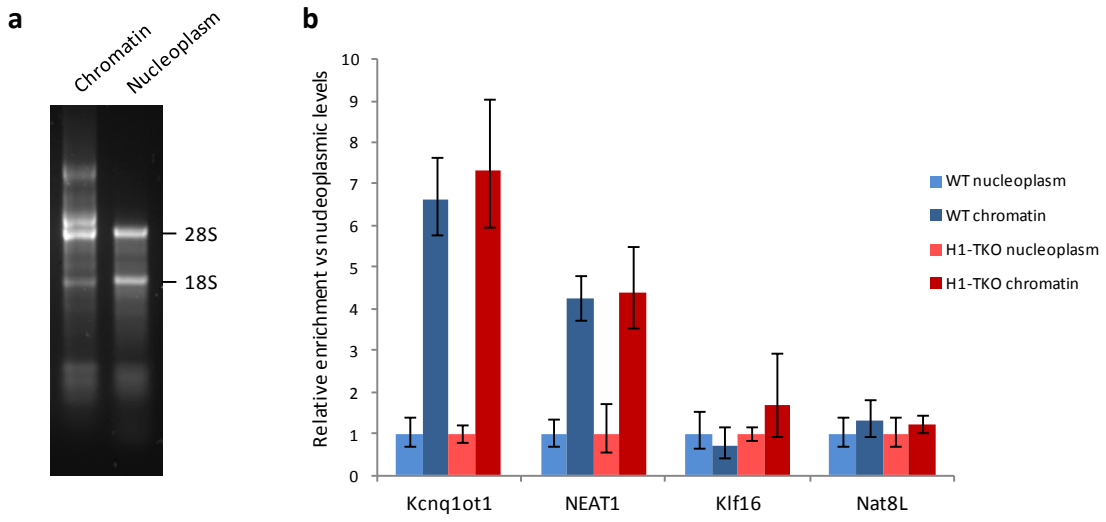


Figure 5: Analysis of chromatin associated RNAs enrichments in cheRNA preparations **(a)** Chromatin and nucleoplasmic RNA fractions run in a non-denaturing 1% agarose gel **(b)** Plot showing the ratio between chromatin and nucleoplasmic levels for four different RNAs. The mean and standard deviation of three biological replicates for WT and H1-TKO mES cells are represented.

Before library preparation, ribosomal RNA was depleted from the samples by a treatment with Ribo-Zero rRNA Removal Kit (Illumina). Libraries were generated using TruSeq Stranded Total RNA Library Prep (Illumina), and sequenced by 1x75 single reads at the Fundación Parque Científico de Madrid.

1.5. RNA polymerase II ChIP-seq

For the extraction and fragmentation of the chromatin, mES cells were crosslinked with 37% formaldehyde during 15 minutes at room temperature, directly added to the culture medium. After stopping the reaction by incubating the samples with 125 mM glycine for 5 minutes, the cells were washed twice with PBS and collected by scrapping in ice-cold PBS supplemented with protease and phosphatase inhibitors (10 μM leupeptin, 100 μM PMSF, 1μM pepstatin, 2 μg/mL aprotinin, 5 mM NaF, 1mM $NaVO_3$). Cells were centrifuged at 200g for 5 minutes, and resuspended in cold Lysis Buffer (50 mM Tris pH 8, 1% SDS, 10 mM EDTA, protease and phosphatase inhibitors), at a concentration of $20 \cdot 10^6$ cells/mL. Afterwards, they were incubated on ice for 20 minutes. The chromatin in the lysate was fragmented with a Covaris sonication system, 40 cycles at 20% intensity, during 20 minutes. 5% of the volume of the lysate was aliquoted, crosslinking was reverted with an overnight incubation with 200 mM NaCl at 65ºC, and the DNA was purified by phenol-chloroform extraction and ethanol precipitation. The concentration of DNA in the aliquot, measured in Nanodrop™ One, was used to quantify the concentration of the chromatin input in the immunoprecipitation.

100 μg of the fragmented chromatin was diluted 1:10 in Dilution Buffer (20 mM Tris pH 8, 1% Triton X-100, 2mM EDTA, 150mM NaCl, protease and phosphatase inhibitors), and 5 μg of human chromatin, obtained from a MCF10A cell line following the same protocol, was added as spike-in control. The mix of the two chromatins was pre-cleared by incubation with 60 μL protein A/G beads during one hour at 4ºC. The beads were removed by centrifugation at 400g 5 minutes, and 25 μg of α-RNApolII antibody (Millipore #05-623) were added. After an overnight incubation at 4ºC with gentle agitation, 200 μL of A/G protein beads were incorporated, followed by a second incubation during 2 hours. The beads were next washed sequentially with four different buffers, all of the supplemented with protease and phosphatase inhibitors: low salt

buffer (10 mM Tris pH 8, 0.1% SDS, 1% Triton X-100, 2mM EDTA, 150 mM NaCl), high salt buffer (20 mM Tris pH 8, 0.1% SDS, 1% Triton X-100, 2mM EDTA, 500 mM NaCl), LiCl buffer (10 mM Tris pH 8, 0.25M LiCl, 1% NP40, 1% Na-deoxycholate, 1mM EDTA) and TE buffer (10 mM Tris pH 8, 1mM EDTA). Finally, the chromatin was eluted with elution buffer (0.1M $NaHCO_3$, 1% SDS), and the DNA was purified with phenol-chloroform extraction and ethanol precipation, after reverting crosslinking as described before.

The libraries were generated following Illumina's recommendations, and were sequenced by 1.75 single-reads in at Fundación Parque Científico de Madrid.

1.6. Mononucleosomal DNA isolation and sequencing

$6 \cdot 10^6$ cells were plated in a p100 tissue culture dish. Another p100 was seeded with the same number of cells as a parallel control of cell number. The following day, the cells on one of the plates were trypsinized and counted in a Neubauer chamber; the cells on the other plate were washed twice with PBS, and crosslinked by a 10 minutes incubation with 10 mL of PBS + 1% formaldehyde at room temperature. The crosslinking reaction was stopped by the addition of 125 mM glycine for 5 minutes. After two washes with PBS, the cells were collected by scrapping in 2 mL ice-cold PBS + protease inhibitors (10 μM leupeptin, 100 μM PMSF, 1μM pepstatin, 2 μg/mL aprotinin). They were collected by centrifugation at 180g for 5 minutes at 4ºC, and resuspended in Homogenization Buffer (10 mM Tris pH 7.4, 1mM EDTA, 0.1mM EGTA, 15 mM NaCl, 50 mM KCl, 0.15mM spermine, 0.5 mM spermidine, 0.2% NP-40, 5% sucrose) at a concentration of $2 \cdot 10^6$ cells/mL. The cells were incubated on ice for 3 minutes to allow the lysis of the cytoplasmic membrane. The nuclei were placed on top of a 3.5 mL sucrose cushion (Homogenization Buffer + 10% sucrose), and centrifuged for 20 minutes at 900g and 4ºC. The nuclei pellet was resuspended in ice-cold Wash Buffer (10mM Tris pH 7.4, 15mM NaCl, 50 mM KCl, 0.15 mM spermine, 0.5 mM spermidine, 8.5% sucrose, 1mM $CaCl_2$) at a concentration of $2 \cdot 10^6$ nuclei/mL, and divided in 1 mL aliquots. Two of these aliquots were treated with 0 and 600 units of MNase (Thermo Fisher #EN0181), for 6 minutes at 25ºC. The enzymatic reaction was stopped by adding 9mM EDTA and 3.5 mM EGTA and placing the samples at 4ºC. The digested samples were stored at -20ºC.

At least three of the not digested aliquots were used to quantify the amount of total genomic DNA per aliquot. Crosslinks were reverted by an overnight incubation at 65ºC in the presence of 150 mM NaCl. Then, the samples were treated with 0.5 mg RNase (DNase free, Roche) for 30 minutes at 37ºC, and with 0.1 mg proteinase K (Roche) overnight at 45ºC. The next day, the genomic DNA was purified by phenol-chloroform extraction and ethanol precipitation, and quantified in Nanodrop™ One.

The mean of the resulting DNA concentrations were used to calculate the amount of the spike-in control added to the digested aliquots: to each aliquot, 37pg of Lma spike-in and 185pg of Taq spike-in per mg of genomic DNA were added. These two spike-in controls are 140-150 bp DNA fragments amplified by PCR from *Leishmania major* and *Thermus aquaticus* genomic DNA (see primers Table S6, Lma-ex and Taq-ex pairs), with no homology to any mouse genomic region. Next, the samples were purified as detailed before, and those treated with 600 units of MNase loaded in a 1.2% agarose gel. The fragments corresponding to the mononucleosomal fraction (around 150 bp) were sliced from the agarose gel and purified using a Speedtools PCR Clean Up Kit (Biotools) following manufacturer's instructions.

For massive sequencing, the libraries were generated from the mononucleosomal fraction and sequenced by 2x75 paired-end reads at Fundación Parque Científico de Madrid.

1.7. Quantitative real time PCR

qPCR were performed in an ABI Prism 7900HT Detection System (Applied Biosystems), using HotStarTaq DNA polymerase (Qiagen) following manufacturer's instructions. For absolute quantification, the Ct of each amplicon was interpolated in a standard curve obtained from the amplification of genomic DNA at five different concentrations (from 0.2ng/μL to 125ng/μL). The primer sequences and the PCR conditions for all the regions amplified are included in primers Table S6. The analyses were carried out with the SDS2.4 software (Applied Biosystems).

1.8. Fiber stretching

10^6 exponentially growing mES cells were pulsed consecutively with 50 mM CldU (Sigma) for 20 minutes and 250 mM IdU (Sigma) for another 20 minutes. Cells were trypsinized and resuspended in cold PBS at a concentration of $0.5 \cdot 10^6$ cells/mL. 2μL of this cell suspension were placed on the top of a microscopy slide, and lysed through the addition of 10 μL spreading buffer (200 mM Tris pH 7.4, 50 mM EDTA, 0.5% SDS) at 30ºC. After 6 minutes of incubation in a humidity chamber at room temperature, DNA fibers were stretched by leaning the slide with a 30º slope. The samples were air dried and fixed with -20ºC 3:1 methanol:acetic acid solution. Next, the slides were incubated with 2.5M HCl solution for 30 minutes at RT, washed three times with PBS, and treated with blocking solution (1% BSA, 1% Triton X-100 in PBS) for 1 hour. Afterwards, the samples were sequentially incubated with the primary antibodies (1:100 Abcam anti-CldU, 1:100 BD anti-IdU and 1:3000 Millipore anti-ssDNA) for one hour, and with the secondary antibodies (1:300 anti-rat IgG AlexaFluor594, anti-mouse IgG1 AlexaFluor488 and anti-mouse IgG2a AlexaFluor647) for another 30 minutes. Finally, the slides were air dried and mounted with Prolong Diamond (Invitrogen). Photographs of the DNA fibers were taken with an Axiovert200 Fluorescence Resonance Energy Transfer microscope (Zeiss) using the 40x oil objective. The images were analyzed with ImageJ software, considering a conversion factor of 1μm=2.59 kb. Three parameters were analyzed: fork rate, measuring the length (in kb) of the IdU track and dividing it by the 20 minutes of the duration of the pulse; the inter-origin distance, measuring the distance between adjacent ORIs (recognized as IdU-CldU-IdU tracks); and fork asymmetry, obtained dividing the length of the two bi-directional IdU tracks rising from a replication origin.

2. Computational methods

2.1. Short nascent strands sequencing

SNS-seq reads were aligned to the mm10 version of the genome using bwa mem algorithm with standard parameters, and filtered with the Samtool view parameter -q 1 to remove multihits. The bedGraph files loaded in the browser were generated with the Bedtool genomecov. The scores of these files were normalized with the total number of aligned reads for each experiment.

To find ORI peaks, the scan-quantile algorithm (Picard *et al.*, 2014) was used, with several own modifications. First, a minor change was introduced, so the software could include in the output the p-value for each individual peak (we later used this p-value as a measure of efficiency). Also, the mathematical formula to obtain the statistic parameter was altered to avoid loss of precision problems, so p-values lower than $1 \cdot 10^{-10}$ could be precisely determined. The scan-quantile algorithm needs a previous segmentation of the genome based on the general coverage level in different regions of the genome. In the original published version, the segmentation was obtained with modified CGH arrays analysis algorithms; we substituted it by a segmentation based on replication timing data from mES (Hiratani *et al.*, 2008), which accurately matches the read coverage differences between regions. After the peak detection, the peaks separated by less than 200 bp were merged (keeping the lowest p-value).

This same peak-calling protocol was applied to a genomic DNA sequencing experiment obtained in parallel experiments by the same method (kindly provided by Dr. Juan Méndez). The peaks which were found in the gDNA were combined in a file, together with sub-telomeric and telomeric regions, as non-mappable regions of the genome. All SNS peaks in these regions were discarded.

Common peaks between biological replicates were obtained with the Bedtool intersect, setting parameters -wa -f 0.1: nonreciprocal, and with at least 10% overlap.

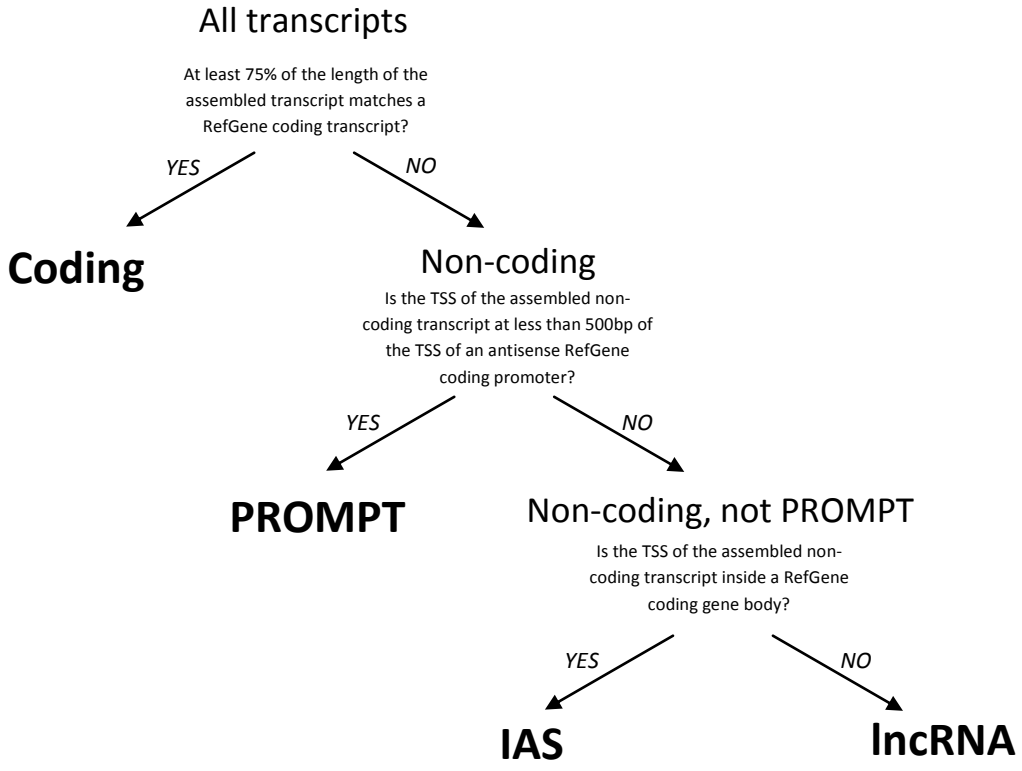
To analyze the genomic distribution of replication origins, the observed percentage of ORI peaks which matched each genomic region was compared with the expected proportion calculated from genomic intervals randomly sampled from throughout the genome. The randomly sampled intervals were obtained with the Bedtool shuffle, parameters -excl non-mappable-regions -noOverlapping. This procedure was repeated 1000 times, and the statistical significance was determined by computing the empirical p-value from the sampling distribution.

Metaplots were generated with the annotatePeaks.pl tool from HOMER suite (Heinz *et al.*, 2010), setting parameters -size 4000 -hist 20: the signal was plotted in 4 kb windows around the center of the peaks, calculated in 200 windows of 20bp each.

2.2. Chromatin enriched RNA sequencing

Reads were aligned to the mm10 reference genome and to the luciferase coding sequence using Tophat2 (Kim *et al.*, 2013) with standard parameters. The bedGraph files loaded in the browser were generated with the Bedtool genomecov. The scores of these files were normalized with the total number of aligned reads for each experiment.

For the transcriptome assembly, reads coming from the six experiments (three WT and three H1-TKO replicates) were pulled, and separated in two files depending on the template strand (Watson or Crick), discriminating them with Samtool view -F 0x10 or -f 0x10, respectively. Spliced reads were discarded from the pull, by removing the entries with a CIGAR string which contained any 'N' character. The remaining reads were used to assemble a "genome-guided" transcriptome with Cufflinks v2.2.1 (Trapnell *et al.*, 2010). This transcriptome was further curated with home-made scripts to remove low abundance transcripts (minimum coverage < 2.5), remove very short transcripts (size < 300 bp), merge proximal transcripts in the same strand (distance < 2.5 kb) and split transcripts which included an already annotated TSS in the RefGene database. These transcripts were classified in four groups: coding, PROMPTs, lncRNAs and internal antisense transcripts (IASs), according to the following diagram:



After the transcriptome assembly, 21702 coding transcripts, 3139 PROMPTs, 12673 lncRNAs and 2904 IASs were detected.

For the differential gene expression analysis, the quantification of reads per transcript was performed with Salmon (Patro *et al.*, 2017) using standard parameters. To select differentially expressed genes, DESeq2 software (Love *et al.*, 2014) was used, setting two different thresholds: adjusted p-value < 0.01 and fold-change > 2.

GO-term enrichment analyses were performed using Panther v14.1 software (Mi *et al.*, 2019). To account for transcription factor and epigenetic enrichments, Enrichr software was employed (Kuleshov *et al.*, 2016). Metaplots were obtained as described in the previous section, separating between Watson and Crick reads for cheRNA-seq experiments as described before.

2.3. Histone H1 knock-down RNA-seq analysis

Reads from total RNA-seq preparations (Izquierdo-Boulstridge *et al.*, 2017) were aligned to the hg19 reference genome using Tophat2 with standard parameters. For the transcriptome assembly, reads coming from the six experiments (two controls, two H1.4-KD and two multi-KD) were pulled, and separated in two files depending on the template strand (Watson or Crick), discriminating them with Samtool view -F 0x10 or -f 0x10, respectively. The reads that matched a RefGene annotated coding gene were removed from the pull, and the remaining reads were used to assemble a “genome-guided” transcriptome with Cufflinks v2.2.1, exclusively from the non-coding part of the genome. This transcriptome was further curated with home-made scripts to remove low abundance transcripts (minimum coverage < 2.5), remove very short transcripts (size < 300 bp), merge proximal transcripts in the same strand (distance < 2.5 kb) and split transcripts which included an already annotated TSS in the RefGene database. Finally, it was merged with the ENSEMBL coding transcriptome using Cuffmerge, and the transcripts were classified in the four same types as before: in the end, 22827 coding transcripts, 2420 PROMPTs, 14843 lncRNAs and 4562 IASs were detected.

The differential gene expression analysis was performed as described in the previous section. In this case, the statistical thresholds were set as adjusted p-value < 0.1 and fold-change > 2.

2.4. RNA polymerase II ChIP-seq

Reads were aligned to mouse mm10 and human hg19 reference genomes using bwa mem algorithm. In addition to the standard total read number normalization, the ratio between mouse and human reads was used to correct the H1-TKO cells metaplot signal, according to this formula:

$$TKOs = TKOr * \frac{mouse\ WT\ reads/human\ WT\ reads}{mouse\ TKO\ reads/human\ TKO\ reads}$$

where

TKO_s is the spike-in normalized RNApolII signal

TKO_r is the total reads normalized RNApolII signal

The meta-gene profile was generated with annotatePeaks.pl tool from HOMER suite (Heinz *et al.*, 2010), setting parameters -size “given” -hist 25: the gene bodies were divided in 25 windows and the number of reads per window was normalized using their length.

2.5. Mononucleosomal DNA sequencing

Paired-end reads were aligned to mm10 reference genome using bwa mem algorithm. Only fragments delimited by two paired reads mapping the same chromosome and separated by less than 250bp were considered for the analysis: those fragments were trimmed 25 bp in both ends and used as input for DANPOS software (Chen *et al.*, 2013).

The metaplots and the analysis of the genomic distribution of differentially positioned or occupied nucleosomes were performed as described in the SNS-seq computational methods section.

RESULTS

1. DNA replication alterations H1-TKO cells

1.1. Analysis of DNA replication initiation landscape

Previous studies in our group had shown that a decrease in the total amount of histone H1 cause major alterations to DNA replication: H1-TKO cells display a decrease in the inter-origin distance, coupled with slow and unstable replication forks (Figure 4). To characterize this further, we first checked if the replicative defects in H1-TKO cells were linked to the activation of extra replication origins, what would be reflected in a hyperphosphorylation of MCM helicases by the CDC7 protein, necessary for DNA replication initiation. We confirmed this increase in MCM phosphorylation by performing a cellular fractionation and testing the resulting cytoplasmic, nucleoplasmic and chromatin protein extracts by western blot, to measure the MCM2 phosphorylated ratio (Figure 6).

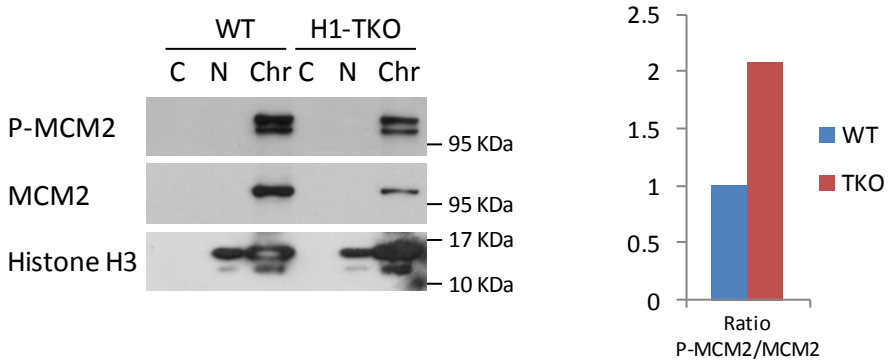


Figure 6: Analysis of MCM2 phosphorylation levels in H1-TKO cells. Left panel; cytoplasmic (C), nucleoplasmic (N), and chromatin (Chr) cellular fractions were assayed by western blot with α-P-MCM2, α-MCM2 and α-H3 specific antibodies. Right panel, the ratio between the intensity of both bands is represented.

Still, the location of these new fired origins was unknown. Our working hypothesis was that the lack of normal amounts of histone H1 could facilitate the binding of pre-RC complexes to chromatin sites which were not accessible in physiological conditions, increasing the number of genomic regions which were licensed during the G1 phase of the cell cycle. We couldn't discard, though, that the instability and stalling of the replication forks was triggering the activation of nearby dormant origins as a compensatory mechanism. In order to discriminate between these two possibilities, and analyze the role of histone H1 in origin selection, we performed

the isolation and sequencing of two replicates of DNA replication intermediates called short nascent strands (SNS) from asynchronously growing H1-TKO cells. This allowed us to generate genomic maps where the SNS enrichment represents a read-out of the probability of a certain region to be a preferential site of DNA replication initiation within the cell population. The detection of the enriched regions or SNS-peaks was done with a modified version of the scan-method peak-calling algorithm described in Picard *et al.*, 2014 (see Materials and Methods and Table S1).

The SNS profile showed a dramatic change in the replication initiation pattern between wild type mES cells and H1 deficient cells. While WT cells showed the typical landscape with well defined peaks, the H1-TKO cells displayed a widespread accumulation of replication intermediates along the genome, what makes the computational detection of origins rather inaccurate: a significant fraction of the peaks are probably false positives, although it is still possible to find certain sites with a slight SNS enrichment (Figure 7a).

Previous work described that mammalian replication origins are preferentially located in certain genomic regions, such as proximal to TSSs and CpG islands. Moreover, those origins that match promoters and CpG islands are the most efficient: they fire in a high percentage of the cells in the population (Sequeira-Mendes *et al.*, 2009; Cayrou *et al.*, 2012; Besnard *et al.*, 2012; Lombraña *et al.*, 2014). To check if this localization pattern is maintained for the H1-TKO SNS enriched sites, we calculated the percentage of origins which were common between the two biological replicates across different genomic regions, and compared that to the percentage expected by chance (Figure 7b). We found that the replication initiation preference for CpG islands and promoters is maintained (or even enhanced) in H1-TKO cells. This is probably due to the genome-wide accumulation of replication intermediates, which restricts reliable origin detection only at the most efficient initiation sites.

Although the few origins we could detect in H1-TKO cells seem to have a similar localization to those identified in the WT situation, the global replication initiation landscape was heavily perturbed (Figure 7a). To measure the extent of this alteration, we plotted the relative firing efficiency of each origin, calculated as the inverse logarithm of the peak-calling algorithm p-value (Figure 7c).

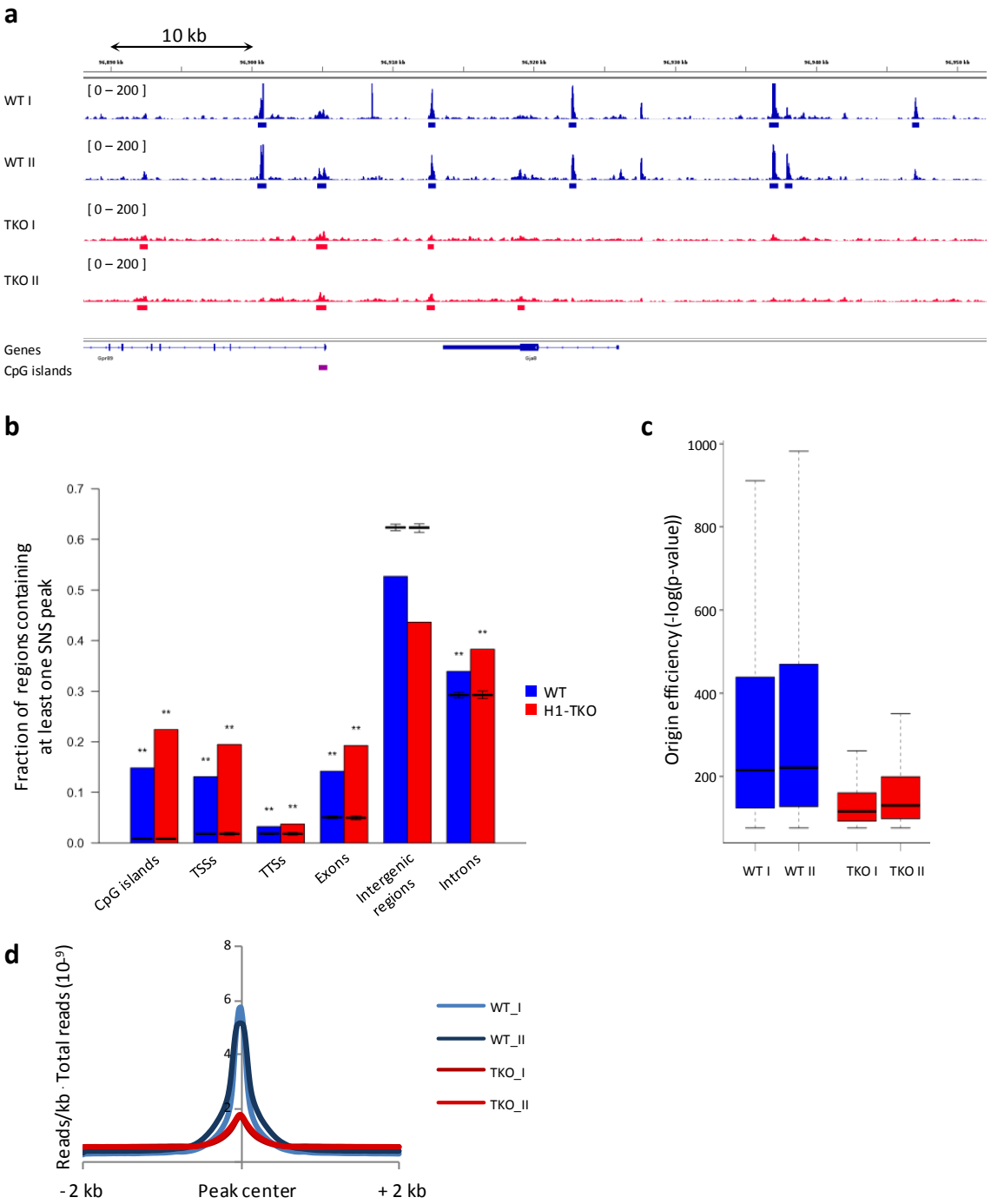


Figure 7: Replication initiation landscape in H1-TKO cells. **(a)** IGV browser screenshot of the SNS-seq coverage in a representative region of chromosome 3. Rectangles below the tracks correspond to the position of the SNS peaks obtained with the scanquantile algorithm. **(b)** Fraction of different genomic regions containing SNS peaks. CpG islands positions were downloaded from UCSC database. TSSs, TTSs, exons, introns and intergenic regions were obtained from GENCODE database. TSSs and TTSs were extended in a 1kb window from the original annotation. The boxplots represent the distribution obtained from 1000 random origin locations. **p-value<0.001 **(c)** Boxplots showing the distribution of efficiencies of SNS-peaks, measured as -log(p-value) **(d)** Metaplot of SNS-seq signal centered in 4kb windows around peak centers.

The efficiency of a peak reflects the probability of the origin to be fired during each S-phase; however, since the peak-calling algorithm uses the signal-to-noise ratio of the peak to calculate this parameter, it can be used also as a measure of the SNS baseline signal in the surrounding area (the higher is the baseline, the relative efficiency gets lower). Besides relative efficiency, we also plotted the SNS signal in a 4kb window around the center of the SNS-peaks (Figure 7d). Together, these analyses illustrate that the relative firing efficiency of replication origins is much lower in H1-TKO cells, likely due to an increase of the baseline signal caused by a widespread accumulation of replication intermediates.

1.2. Characterization of the alterations in DNA replication initiation

In previous studies in our lab, we found that the replication perturbation in H1-TKO cells was dependent on active transcription: the slowdown and the asymmetry of the forks were reversed through the addition of drugs that inhibit the activity of the RNA polymerase II, like DRB or α-amanitin (Figure 4). One possible explanation for these results, which could also account for the presence of replication intermediates genome-wide, would be the stalling of the replication forks due to collisions with the transcription machinery. These transcription-replication conflicts would trigger the activation of dormant origins, but could also require the DNA-polymerase complexes to re-prime ahead of the transcriptional block. In both cases, new short leading strands with an RNA primer would be generated, what would cause the increase in the overall genome-wide initiation signal described in the previous section.

If the altered replication landscape in H1-TKO cells is due to conflicts with RNA-polymerase II active complexes, it could be predicted that the accumulation of replication intermediates would not be homogeneous along the genome, but preferential at transcribed genes relative to non-transcribed genomic regions. To check this, we plotted the relative efficiency of replication origins both in genic and non-genic regions: if there is an increase of the background signal in genic regions, this should be reflected in a drop in the efficiency. However, this does not occur: the origin efficiency inside genes is higher, both in WT and H1-TKO conditions (Figure 8).

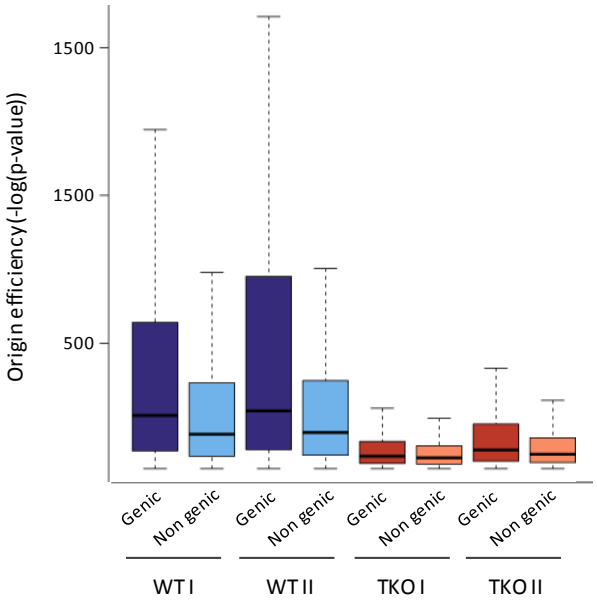


Figure 8: Analysis of peak effiencies inside genic and non genic regions, measured as the -log(p-value) of the peak. Genic regions were obtained from GENCODE database: they include UTRs, exons and introns of annotated genes.

In this analysis, we only took into account annotated genes. Recently, it has been reported that more than 85% of the genome is being transcribed (Hangauer *et al.*, 2013), but many of the RNAs that are generated are unstable, with very short lifetimes, so the majority of these transcribed regions are not included in gene databases. It was therefore possible that the replication alterations in H1-TKO cells were not caused by mRNA transcription, but by non-productive transcription arising from non-annotated genes, which would be present in most of the genome. To address this possibility, we designed the experiments described in the next chapter.

2. Characterization of differentially expressed transcripts in histone H1 defective cells

2.1. Differential chromatin-bound RNA analysis

To enrich in non-coding and unstable transcripts, we decided to isolate the chromatin-bound RNA fraction (cheRNA, see Materials and Methods) of WT and H1-TKO cells. This technique has been shown to preferentially detect RNA that is bound to chromatin through a RNA-polymerase molecule (Werner and Ruthenburg, 2015), thus

enabling to analyze partially processed transcripts, attached to chromatin by either an active or stalled RNA polymerase molecule, which could pose an obstacle to the DNA replication machinery. The inclusion of a spike-in luciferase RNA molecule allowed us to assess potential changes in the overall amount of chromatin-bound RNA.

We performed the isolation and sequencing of cheRNA from three independent replicates of WT mES cells and H1-TKO cells. We started by quantifying the percentage of luciferase spike-in generated reads versus total number of reads for each replicate, which can be used as a measure of the quantity of chromatin-bound RNA per cell (Table S2). No statistically significant difference was found between the two conditions (p-value > 0.05), indicating that the lack of histone H1 does not cause a global alteration in the amount of chromatin-bound RNA, or at least, detectable by this mean (Figure 9a).

In order to identify alterations in specific transcripts, we designed a computational pipeline which includes the *de novo* assembly of a transcriptome and the classification of the transcripts in four major categories: coding transcripts, PROMPTs, internal antisense transcripts (IASs) and long intergenic non-coding RNAs (lncRNAs) (see Materials and Methods for details). Surprisingly, in contrast with previous steady-state transcriptomic analysis, we found a high number of differentially expressed transcripts from the four categories, (Figure 9b and 9c), in spite of setting very strict thresholds for the statistical detection (fold-change > 2 and adjusted p-value < 0.01).

Besides the number of differentially expressed transcripts, the proportion between WT overexpressed genes and TKO overexpressed transcripts for each category was analyzed. Regarding coding transcription, the lack of histone H1 causes an up-regulation of around 1300 transcripts, roughly a similar number than those that are down-regulated. In contrast, the three non-coding categories are mostly up-regulated in H1-TKO cells: there are more than three times more upregulated IASs, PROMPTs and lncRNAs than in WT cells (Figure 9a, 9b and 9c). This indicates that the reduction in histone H1 content is causing a widespread accumulation of non-coding transcripts in chromatin, either because of an enhanced RNA polymerase II activity, or an increased stability and residence time (or both).

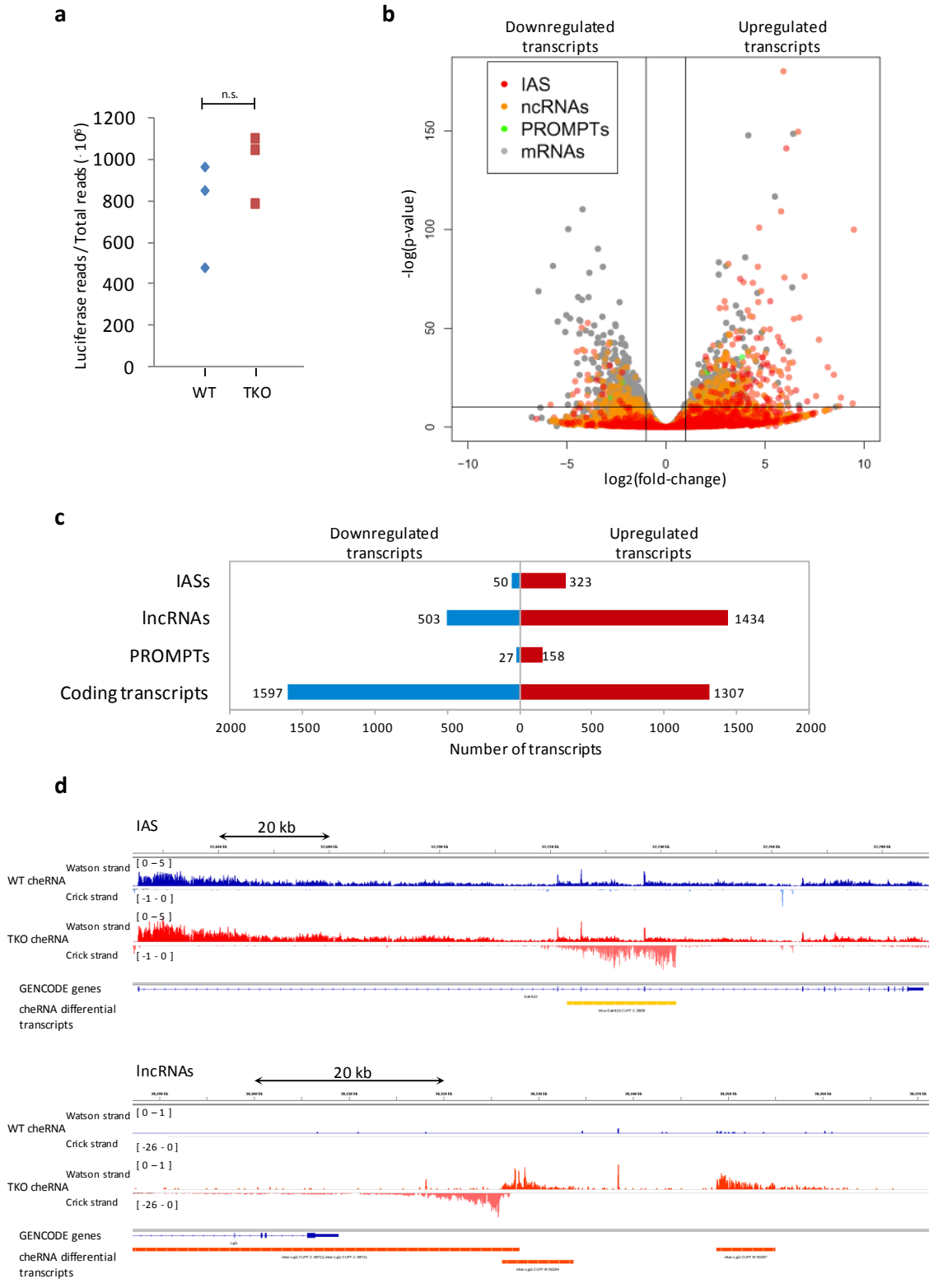


Figure 9: Differential expression analysis between WT and H1-TKO cheRNA-seq. **(a)** Percentage of Luciferase reads in WT and H1-TKO replicates. *n.s.=not significant, t-test p-value=0.30. **(b)** Vulcano plot showing the log(fold-change) and the -log(p-value) for each transcript **(c)** Number of differentially expressed transcripts between WT and H1-TKO cells for each category **(d)** Representative IGV browser snapshots of transcripts specifically accumulated in the chromatin of H1-TKO cells. Upper panel, an internal antisense transcript for *Galnt10* gene. Lower panel, several lncRNAs adjacent to the *Lgi1* gene.

Even with the high number of differentially accumulated non-coding transcripts in H1-TKO cells chromatin, it was difficult to explain the altered replication initiation pattern, which is perturbed genome-wide and not only around these sites of cryptic transcription. However, there are still many cheRNA reads that are distributed all along the genome, but didn't reach the minimum statistical threshold level to be included in a *de novo* assembled transcript (Figure 10a). When we calculated the number of reads that have stayed out of the transcriptome assembly, we could detect that this transcriptional noise is higher when the levels of histone H1 are reduced (Figure 10b). This means that there are still many low level transcripts that we couldn't analyze, but which are also upregulated in H1-TKO cells, and could potentially interfere with DNA replication.

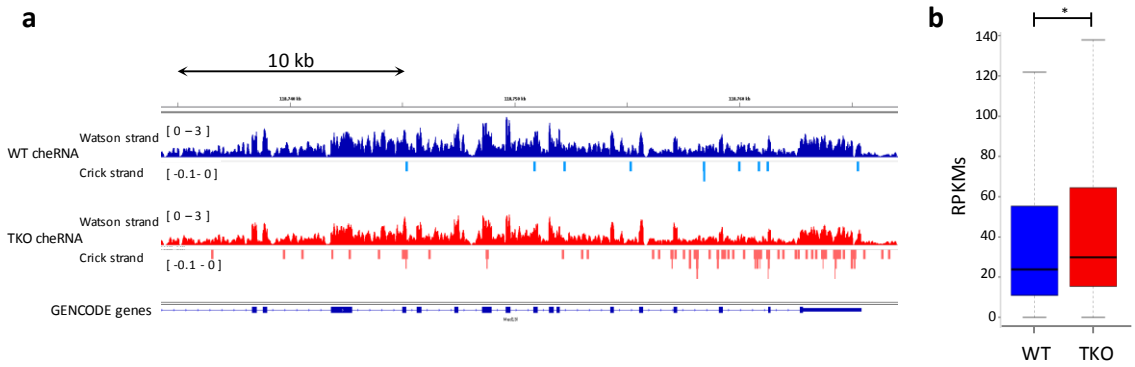


Figure 10: Analysis of background reads not statistically included in de novo transcriptome **(a)** Representative IGV browser snapshot of a region not considered as a transcript with differential cheRNA reads accumulation in H1-TKO cells (in the Crick strand) **(b)** Boxplot showing the distribution of RPKMs in intergenic regions in WT and H1-TKO cells. *Mann-Whitney-Wilcoxon test p-value < $2.2 \cdot 10^{-16}$

2.2. Characterization of differentially expressed coding cheRNAs

Previous transcriptomic analysis concluded that the differentially expressed transcripts in H1-TKO cells were few, and did not belong to any specific gene category, besides certain enrichment in imprinted genes (Fan *et al.*, 2005; Geeven *et al.*, 2015). Conversely, in our chromatin enriched preparations there is an important number of coding transcripts that are differentially enriched or depleted. In order to further characterize this group of differential coding genes, we performed a GO-term enrichment analysis, using Panther software with standard parameters. We detected several enriched categories, most of them related to cell differentiation and development (Figure 11a). Interestingly, another category called "viral process" was also statistically over-represented; the reason is the activation in H1-TKO cells of the

OAS/RNaseL pathway, involved in the detection and degradation of dsRNA. To note, transcriptional activation of the interferon responsive genes was previously reported in human breast cancer cells conditionally knocked-down for the three orthologous histone H1 variants (Izquierdo-Boulstridge *et al.*, 2017).

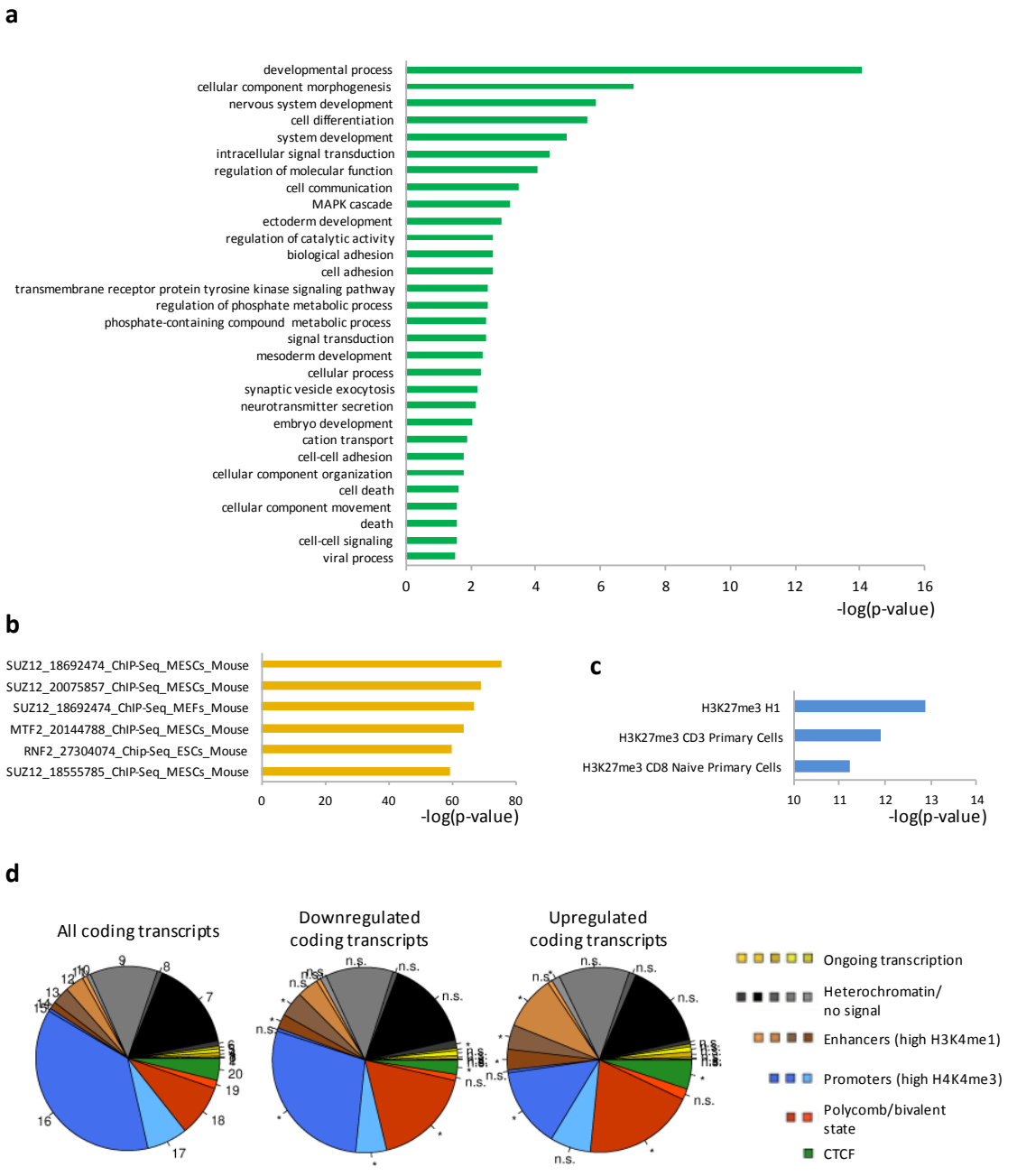


Figure 11: Characterization of differentially accumulated coding transcripts. **(a)** Panther GO-term enrichment analysis **(b)** Enrichr analysis of overrepresented transcription factor binding sites at the promoters of differential coding transcripts. The -log(p-value) for each ChIP-seq database entry is plotted. **(c)** Enrichr analysis of overrepresented epigenetic marks in the promoters of differential coding transcripts. The -log(p-value) for each ChIP-seq database entry is plotted **(d)** Pie plots showing the percentage of transcripts whose promoter matches each chromatin state. The percentage was compared with the expected percentage obtained from 100 random permutations of the differential transcripts, and the p-value was calculated. *p-value<0.01. Chromatin states are from Juan *et al.* (2016).

We next analyzed the epigenetic state of the promoters of these genes by using the Enrichr software to look for any specific epigenetic feature or transcription factor binding sequence that was enriched at the promoters of this set of genes. As expected for developmental and differentiation related genes, their promoters were significantly enriched for Polycomb features: SUZ12 binding sites and H3K27 trimethylation (Figure 11b and 11c).

To gain further insight in this epigenetic configuration, we crossed the promoter location of the differentially expressed genes with the 20 chromatin states described for mouse embrionary stem cells in Juan *et al.*, 2016. This study was based on ChIP-seq studies of multiple histone modifications and chromatin-binding proteins; in the end, the authors segment the genome in 20 different states which share similar epigenetic features. In addition to an enrichment in the Polycomb states, we also found that the differentially expressed transcripts are preferentially synthesized from promoters with mono-methylated H3K4 (H3K4me1, belonging to the “enhancer” category). This enrichment for H3K4me1 promoters is exclusive of genes that are enriched in H1-TKO cells, and is absent in the case of the depleted genes (Figure 11d).

One possible explanation for these differences on the levels of coding transcripts between WT and H1-TKO cells could be a different epigenetic configuration of their promoters. To test this hypothesis, we made use of published ChIP-seq data of a range of epigenetic features performed in the same cells (Geeven *et al.*, 2015). In this study, the authors found that the lack of histone H1 was correlated with changes in the level of H3K4 mono and trimethylation at different regions of the genome; our aim was to assess if these regions matched the promoters of our differentially expressed coding genes. Since we found this enrichment in Polycomb promoters in our set of transcripts, the H3K27me3 mark was also included in the analysis. We performed metaplot profiles for these epigenetic marks and the cheRNA signal in a 4kb window around the TSS of the differentially expressed transcripts (Figure 12a). This analysis led us to the conclusion that there were no clear differences in the epigenetic configuration of the promoters of these set of genes between WT and H1-TKO cells. There is a change in H3K4me3 and H3K27me3 between WT and H1-TKO conditions, but it is not specific from differentially expressed promoters, as is also present when including all the TSS in the analysis (Figure 12a, left graph). Rather than a biological explanation, it seems more probable that this difference was due to slight variations of

the quality of the ChIP-seq data. Still, the enrichment in H3K27me3 at the promoters of differential transcripts could be confirmed.

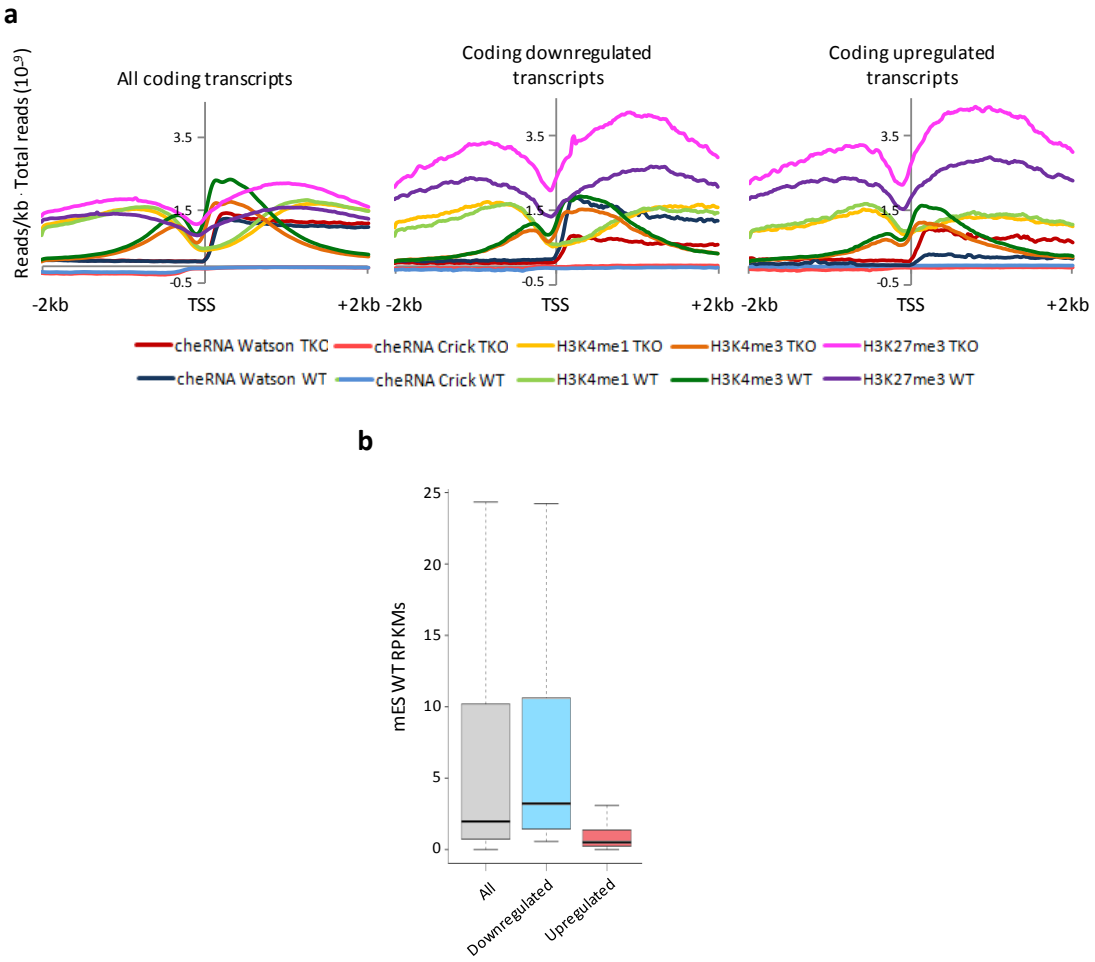


Figure 12: Analysis of the epigenetic configuration and overall levels of differentially accumulated coding transcripts. **(a)** Metaplots of the indicated ChIP-seq signals and cheRNA-seq reads for WT and H1-TKO cells, plotted in 4kb windows around TSSs. H3K4me1 WT and H1-TKO signals were multiplied by a scale factor of 2, to facilitate the visualization in a single plot **(b)** Boxplots showing the distribution of the RPKMs of differential coding transcripts in WT mES cells.

While differences in epigenetic marks between enriched and depleted genes couldn't be detected in the TSS neighborhood, we found a clear change regarding expression levels. Up-regulated genes were not expressed at high levels in H1-TKO cells, but they barely reach the minimum detection level in WT cells. This does not happen for down-regulated transcripts: their expression level was not so different from the mean level of all transcripts. This finding was confirmed by measuring the RPKMs (reads per kilobase per million mapped reads) of the differentially expressed genes in WT mES cells (Figure 12b): up-regulated genes in H1-TKO cells were very lowly expressed in WT conditions.

2.3. Characterization of differentially expressed chromatin associated lncRNAs

Non-coding transcripts annotated *de novo* from cheRNA preparations are not easy to characterize, since they are generally not conserved, and most of them do not match an already annotated gene in public genomic databases: only 24.5% of them correspond with an ENSEMBL non-coding transcript. Previous studies (Werner *et al.*, 2017; Werner and Ruthenburg, 2015) showed that chromatin associated non-coding RNAs could work as “enhancer RNAs”, colocalizing with proximal promoters and increasing their activity. In a genome-wide approach, the presence of a chromatin associated RNA near a promoter is considered a good predictor of its transcriptional activity, even at a greater extent than the presence of traditional enhancer marks (H3K27ac/H3K4me1). To confirm that the lncRNAs that had been detected in our cheRNA isolations were enhancing the activity of proximal promoters, we divided all the coding genes in seven categories, depending on their proximity to any lncRNA, and calculated their transcription level (in RPKMs). From this analysis we could conclude that the presence of a lncRNA near a coding gene increase its transcriptional activity (Figure 13a).

In agreement with this interpretation, when projecting the location of lncRNA promoters with the 20 chromatin states of Juan *et al.* 2016, we found that a large fraction of the non-coding transcripts are generated from genomic regions marked by H3K4me1, traditionally recognized as enhancer regions (Figure 13b, left graph). Interestingly, this preferential location of non-coding promoters in H3K4me1 sites was even more pronounced for the non-coding transcripts that are up-regulated in H1-TKO cells (Figure 13b, right graph).

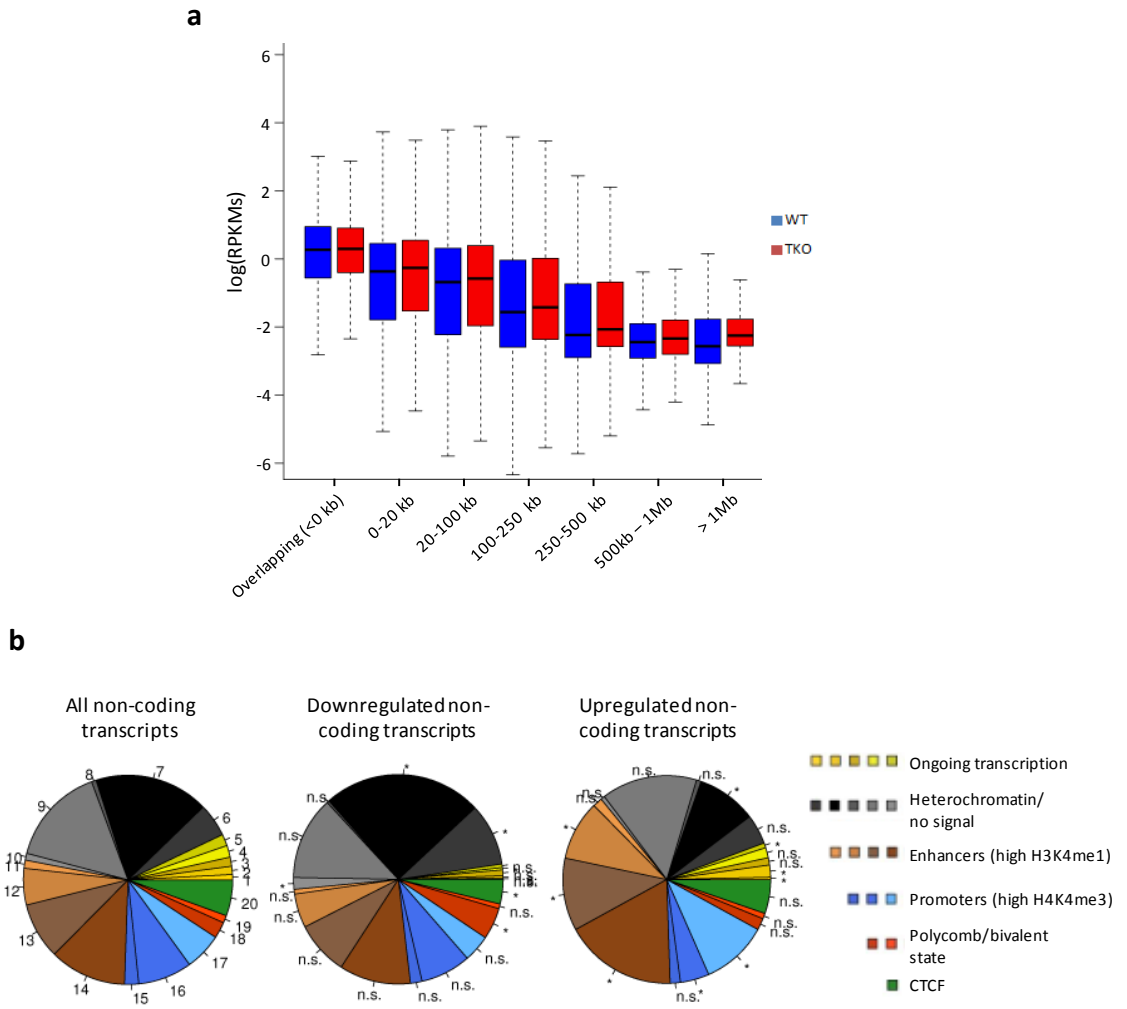


Figure 13: Characterization of differentially accumulated long non-coding transcripts **(a)** Boxplots of the distribution of log(RPKMs) of coding genes located at several distances of a lncRNA, both in WT and H1-TKO cells **(b)** Pie plots showing the percentage of lncRNAs whose promoter matches each chromatin state. The percentage was compared with the expected percentage obtained from 100 random permutations of the differential transcripts, and the p-value was calculated. *p-value<0.01

Next, we decided to study in which cellular processes the differentially expressed lncRNAs could play a role. As mentioned before, the absence of these lncRNAs from any genomic database prevented us from knowing their precise function. However, the fact that they could be regulating the transcriptional activity of proximal genes allowed us to carry out an indirect analysis. For each differentially expressed non-coding transcript, the nearest coding gene was taken; this list of genes was used to perform a GO-term enrichment analysis (Figure 14a). We found that the lncRNAs that show a differential enrichment in H1-TKO cells are preferentially located near

genes involved in development and RNA polymerase II transcription. A deeper analysis of this last group revealed that it was integrated by a set of transcription factors whose expression is modulated during cell differentiation and development, including *Sox2*, *Nanog*, *Oct3/4* and *c-Myc*, four major determinants of the pluripotency state of embrionary stem cells.

Taking into account that the differentially expressed coding genes were enriched in very similar GO-terms than the set of genes proximal to a differential lncRNA (see point 2.2), we checked if the accumulation of non-coding transcripts in H1-TKO cells was altering the expression of the proximal coding genes. For each differentially expressed lncRNA, we found the nearest coding gene, and calculated the fold-change between WT and H1-TKO conditions for both of them. When plotting the fold-change of lncRNAs versus the fold-change of their potential coding targets, we could evaluate if both parameters were correlated (Figure 14b). For down-regulated lncRNAs, the decreased level of the non-coding transcripts was coupled with a clear reduction of the transcriptional activity of their proximal coding genes. Surprisingly, this did not happen for up-regulated lncRNAs: most of their targets did not show any substantial alteration in their transcriptional level. This suggests that the accumulation of non-coding transcripts in the chromatin of H1-TKO cells was abnormal, in the sense that it did not cause the transcriptional response that it should trigger in physiological conditions: enhancer RNAs are upregulated, but seems that they were not carrying out their normal function.

When looking at the upregulated lncRNAs that were proximal to the pluripotency related genes *Sox2*, *Nanog* and *Oct3/4*, we could confirm that they were enhancer RNAs: their transcription start sites matched the superenhancer regions that regulate these three genes in embrionary stem cells (Li *et al.*, 2014; Blinka *et al.*, 2016; Liao *et al.*, 2013). The three of them constituted good examples of the upregulation of an enhancer RNA which does not trigger any change in their coding target transcriptional activity: *Sox2*, *Nanog* and *Oct3/4* are not differentially expressed in H1-TKO cells (Figure 14c, upper panel). This contrasts with the situation of the *c-Myc* gene: the downregulation of its superenhancer RNA (Dave *et al.*, 2017) does cause a reduction of the coding transcript level (Figure 14c, lower panel).

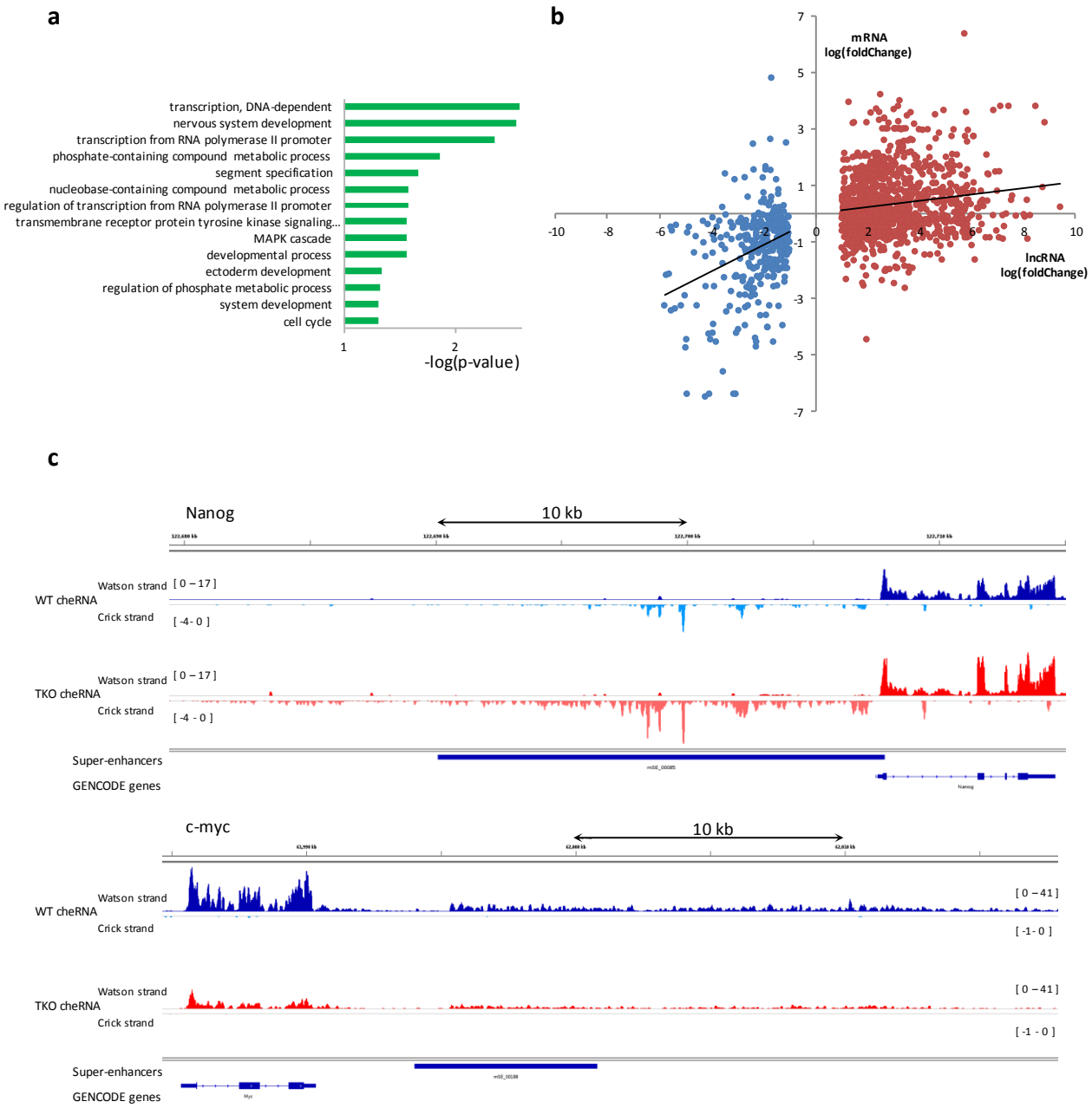


Figure 14: Analysis of the effect of differentially accumulated lncRNAs on proximal coding genes **(a)** Panther GO-term enrichment of the set of genes which are near a differential lncRNA, -log(p-value) for each term is plotted **(b)** Plot showing the correlation between the fold-change of a lncRNA between WT and H1-TKO cells (X-axis) and the fold-change of the nearby coding gene (Y-axis) **(c)** IGV browser snapshots of the eRNAs that regulate Nanog (up) and c-myc (down). The upregulation of *Nanog* eRNA does not have an effect on Nanog expression, but the downregulation of *c-myc* eRNA does trigger the decrease of c-myc transcription.

As we had previously done with differential coding transcripts, we decided to investigate whether the changes in the levels of lncRNAs in H1-TKO cells were due to an alteration of the epigenetic state of their promoters. The ChIP-seq signal of several chromatin marks (H3K4me1, H3K4me3 and K3K27me3, Geeven *et al.*, 2015) was plotted in 4kb windows around the TSS of differentially expressed non-coding transcripts, for both WT and H1-TKO cells (Figure 15). Once again, we did not appreciate any clear difference in the epigenetic configuration of the promoters of these set of genes between the two conditions. Though, the enrichment in H3K4me1 around the TSS of the upregulated lncRNAs could be confirmed.

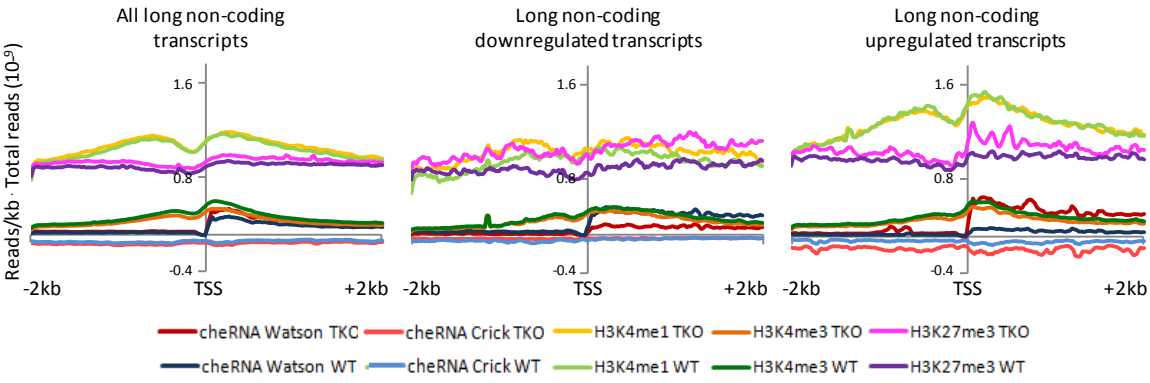


Figure 15: Metaplots of epigenetic marks ChIP-seq signal for WT and H1-TKO cells, plotted in 4kb windows around the TSSs of lncRNAs. H3K4me1 WT and TKO signals were multiplied by a scale factor of 2 in the same manner than coding transcript metaplots, to allow the direct comparison between them.

2.4. Characterization of differentially expressed IASs and PROMPTs

PROMPTs and IASs have been shown to regulate the transcriptional activity of their coding partner genes in different ways, acting like activators or repressors (Preker *et al.*, 2011; Lloret-Llinares *et al.*, 2016; reviewed in Wight and Werner 2013). Taking into account the accumulation of these types on non-coding transcripts in H1-TKO cells (Figure 9), we investigated if this had any effect on their neighboring genes. To check that, we performed a similar analysis we had done for lncRNAs coding targets, plotting the fold change of the PROMPT (or IAS) versus the fold change of its coding partner (Figure 16a and 16b). We could detect a similar trend to that unveiled for lncRNAs: while a decrease in the expression of PROMPTs and IASs is correlated with a lower expression of the corresponding coding gene, the overexpression of these ncRNAs seems to have no effect. This suggests that the accumulation of these types of non

coding transcripts was pathological, and it is not triggering the normal transcriptional regulatory effects on nearby genes that it would cause in physiological conditions.

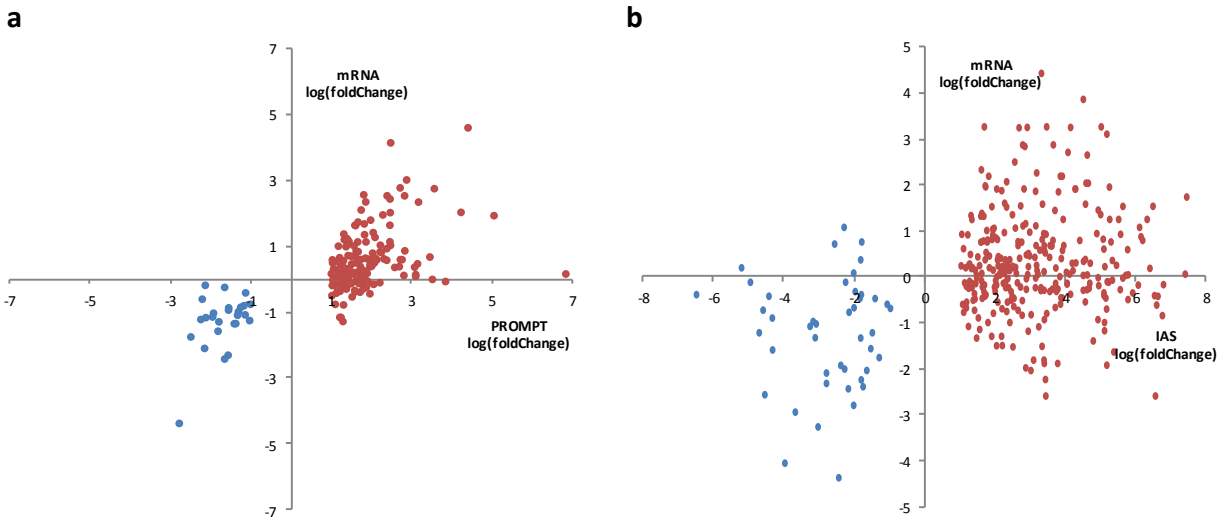


Figure 16: Plots showing the correlation between the fold-change of PROMPTS **(a)** and IASs **(b)** between WT and H1-TKO cells (X-axis) and the fold-change of the corresponding coding gene (Y-axis)

2.5. Differential RNA-seq analysis in histone H1 inducible knock-downs

From the transcriptomic analysis of cheRNAs, we concluded that there was a widespread accumulation of lncRNAs in the chromatin of H1-TKO cells. One of the phenotypes that had been previously described in these cells was their inability to differentiate, likely due to the permanent activation of several pluripotency markers, including *Nanog* and *Oct3/4* (Zhang *et al.*, 2012). During pluripotent stem cell differentiation, the transcription of many lncRNAs is modulated, since they have different roles in the regulation of superenhancer activity, genomic imprinting or Polycomb recruitment (reviewed in Lee *et al.*, 2017). To confirm that the accumulation of lncRNAs in chromatin was a direct consequence of histone H1 reduction, instead of an indirect factor related to the perturbation of differentiation, we studied the transcriptional state of a different cellular model with reduced amounts of histone H1 in an inducible manner. This model is a human T47D breast cancer cell line, which has been stably transfected with several doxycyclin-inducible shRNAs that target individual variants of the histone H1 (Izquierdo-Boulstridge *et al.*, 2017). Specifically, we reanalyzed published total RNA-seq data from two of these cell lines: the individual knock-down of H1e variant (sh-H1e), and a combined knock-down for variants H1c, H1d and H1e (sh-multiH1).

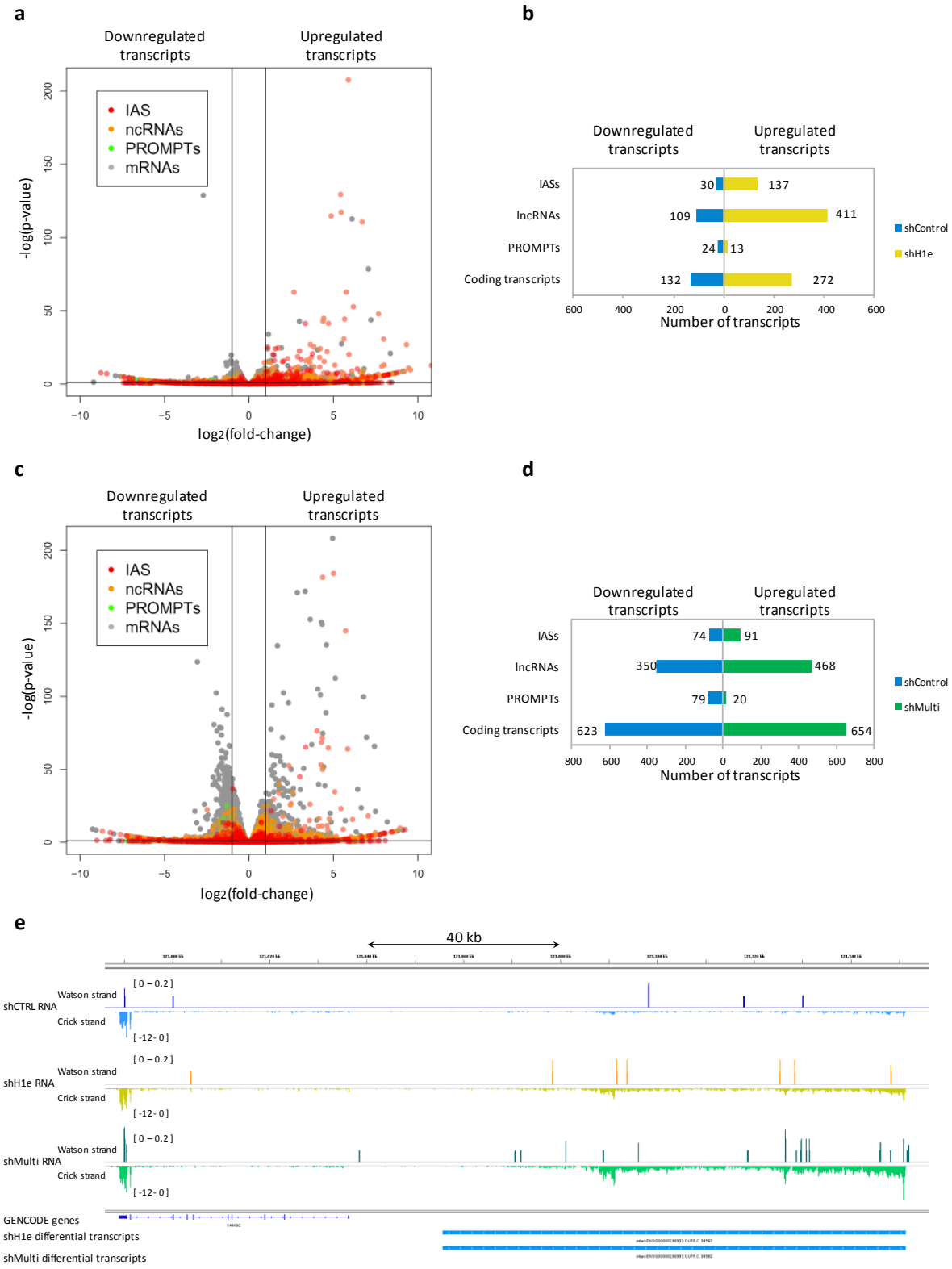


Figure 17: Differential expression analysis in shH1e and shMultiH1 TD47 cells. **(a)** Vulcano plot showing the log(fold-change) and the -log(p-value) for each transcript in shH1e cells. **(b)** Number of differentially expressed transcripts between shCTRL and shH1e cells for each category. **(c)** Vulcano plot showing the log(fold-change) and the -log(p-value) for each transcript in shMultiH1 cells. **(d)** Number of differentially expressed transcripts between shCTRL and shMultiH1 cells for each category. **(e)** Representative IGV browser snapshot of a differentially expressed lncRNA, present in both shH1e and shMulti cells. Total RNA-seq data are from Izquierdo-Boulstridge *et al*. (2017).

We applied the same computational pipeline designed for chromatin enriched RNAs with minor modifications (see Materials and Methods), including the *de novo* assembly of a transcriptome. Analyzing total RNA preparations instead of chromatin enriched RNA had several limitations regarding non-coding transcription. First of all, it was not possible to assemble many non-coding transcripts, apart from those that were already annotated, since they are preferentially enriched when isolating cheRNAs: up to 72.6% of them were present in the ENSEMBL non-coding database. Moreover, their low relative amounts made more difficult to find significant differences between the knock-down and the WT conditions. Regardless these differences in RNA preparations and computational analysis, the general picture was highly reminiscent to that unveiled in H1-TKO cells: non coding transcription was specifically enhanced both in the sh-H1e and the sh-multiH1 cells upon doxycyclin induction (Figure 17).

3. Transcriptional alterations in H1-TKO cells

3.1. RNApolII ChIP-seq analysis

To check if the non-coding RNAs detected in the transcriptome assembly were associated to transcriptional complexes, we performed RNApolII ChIP-seq experiments in WT and H1-TKO mES. During the initial steps of the immunoprecipitation, we added the same amount of human chromatin per cell number to both inputs as a spike-in control, so we could detect quantitative differences in the total amount of chromatin-bound RNApolII between both cell types (the antibody is able to recognize both the mouse and the human protein, see Materials and Methods). After sequencing, the ratio between the number of reads coming from the mouse genome and the reads coming from the human genome was used to calculate the difference in total RNApolII between the two conditions (Table S3 and Materials and Methods). As shown in Figure 18a, this analysis suggests that there was a 12.3% more chromatin-bound polymerase in H1-TKO cells. To note, these results are derived from a single ChIP experiment, a biological replicate is being currently analysed.

Next, we sought to determine where this excess of RNApolII was located in the genome. If the reduction of histone H1 levels in chromatin facilitates the recruitment of the transcription machinery at promoters, it could be expected that these 12% excess of RNApolII was evenly distributed around the transcription start sites (TSS) and bodies

of the genes. However, taking into account the accumulation of non-coding transcripts in histone H1-depleted cells, another likely possibility will be to find RNApolII enrichment associated to these specific transcripts. In a first approach to answer this question, we decided to map RNApolII ChIP-seq reads to the mouse genome, and quantify the number of reads coming from coding gene bodies, promoters, and intergenic regions (Figure 18b), obtained from the Gencode database. We found that, in H1-TKO cells, the percentage of reads coming from intergenic regions was slightly higher, suggesting that the excess of RNApolII is not uniformly distributed along the genome.

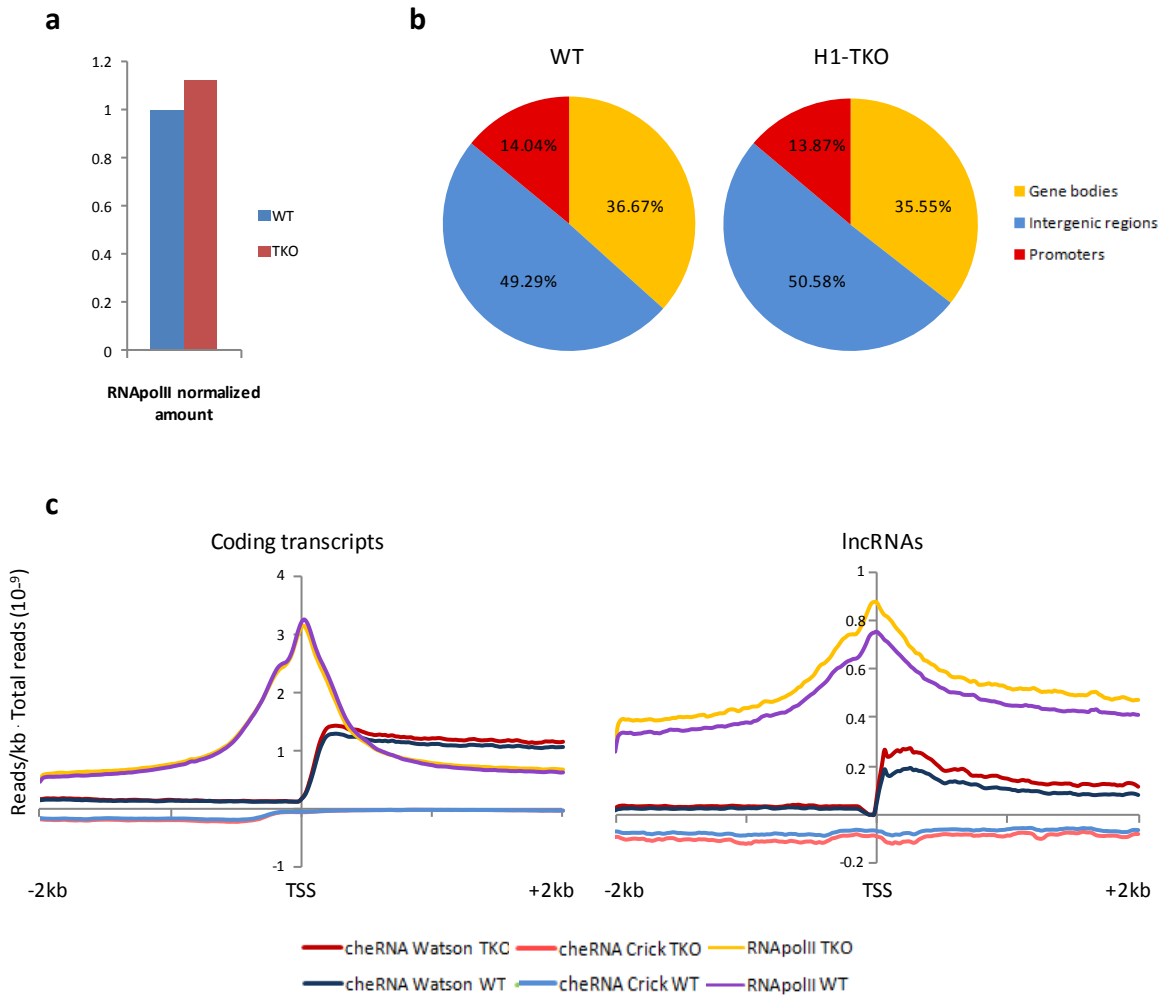


Figure 18: Analysis of the location of the increased amount of RNA polymerase II in the chromatin of H1-TKO cells **(a)** Normalized amounts of immunoprecipitated RNApolII molecules in WT and H1-TKO cells **(b)** Pie plots showing the distribution of RNApolII reads in different genomic regions. Promoter annotation was downloaded from GENCODE coding transcripts database, as the locus comprised from TSS-1kb to TSS+1kb; gene bodies encompass the region from TSS+1kb to TTS; intergenic regions are the remainder genome. **(c)** Metaplots of cheRNA-seq and spike-in normalized RNApolII ChIP-seq signals, plotted in a 4kb window around the TSS. Note that the scale of the Y-axis is not the same in both plots, in order to facilitate the visualization of RNApolII signal differences between WT and H1-TKO conditions. RNApolII WT and TKO signal was multiplied by a scale factor of 1:2, to facilitate the visualization in a single plot.

To confirm this observation, we plotted the spike-in normalized RNApolII signal in 4kb windows around coding and non-coding TSS, obtained from the cheRNA assembled transcriptome (Figure 18c). While we found almost no difference regarding coding promoters, non-coding promoters showed a slightly increased occupancy of RNA polymerase II along the whole analyzed region, confirming that the total levels of RNApolII are higher in H1-TKO cells chromatin.

We next plotted the spike-in normalized RNApolII signal in 4 kb windows around the TSS of differentially expressed coding transcripts, lncRNAs and IASs (Figure 19). These analyses showed that a change in the amount of any type of RNA is accompanied by a parallel change in the levels of RNApolII at their TSS-proximal region. Thus, we concluded that the differentially expressed RNAs we have detected were actually attached to chromatin by a molecule of RNApolII. This does not mean, however, that they were being actively transcribed: although associated to a nascent RNA, the transcription machinery could be stalled or paused.

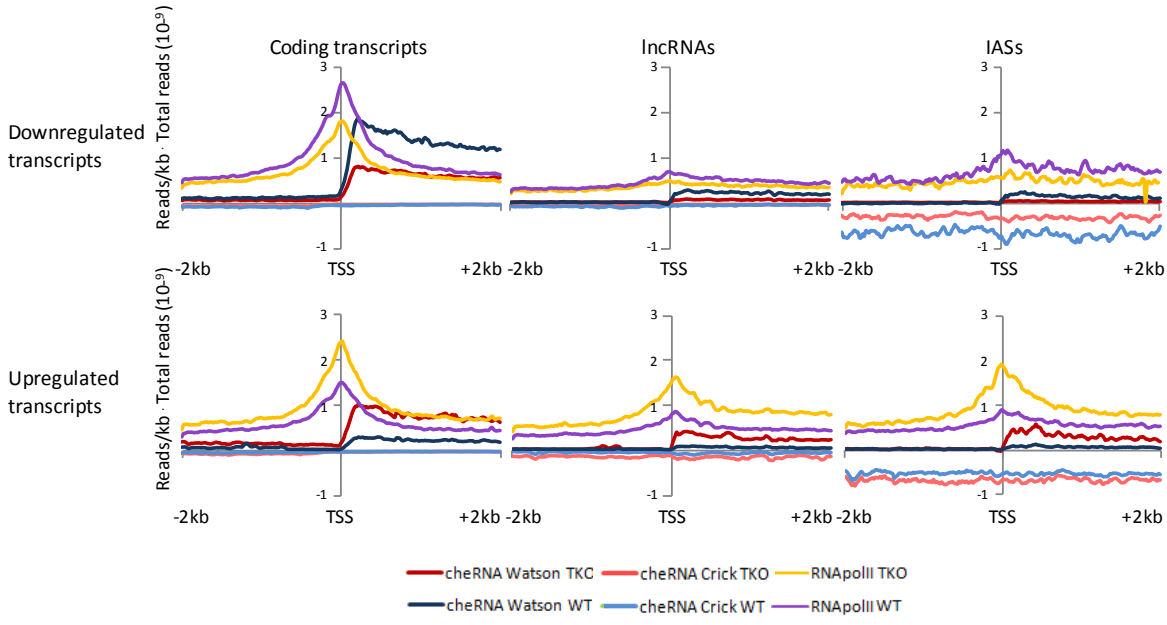


Figure 19: Metaplots of spike-in normalized RNApolII signal and cheRNA-seq at the promoters of the different categories of transcripts, plotted in a 4kb window around the TSS. RNApolII WT and TKO signal was multiplied by a scale factor of 1:2, to facilitate the visualization in a single plot.

In these plots, a qualitative difference between down-regulated and up-regulated ncRNAs could be seen. The promoters of lncRNAs and IASs that were enriched in WT conditions recruited more RNApolII in WT cells than in H1-TKO cells, as expected; however, the level of RNApolII in these promoters was still much lower than the general level at coding promoters (Figure 19, upper row). Conversely, the promoters of lncRNAs and IASs that were enriched in H1-TKO cells recruit almost as much RNApolII as a coding promoter (Figure 19, lower row). These results are in agreement with the presence of increased levels of RNApolII molecules along intergenic regions as possible barriers for the passage of replication forks, likely enhancing transcription-replication conflicts.

3.2. Promoter pausing alterations

Besides differential RNA polymerase II recruitment at promoters, we also analyzed whether other steps of the transcription cycle were altered in H1-TKO cells, such as promoter pausing release, elongation and termination. Chromatin configuration and nucleosomal stability regulate the promoter pausing release or the RNA polymerase II (Jimeno-González *et al.*, 2015a; Weber *et al.*, 2014), so a reduction of the levels of histone H1 around the TSS of certain genes could impair transcriptional regulation at this stage of the cycle. To check if promoter pausing was altered, we calculated the RNApolII pausing index of coding genes in WT and H1-TKO cells (Figure 20a). The pausing index was defined as the ratio between the ChIP-seq RPKMs at the promoter (from TSS-500bp to TSS+500bp) and the gene body (from TSS+500bp to TTS). This index is higher when RNA polymerase II suffers a longer delay before engaging into productive elongation. We could detect a small decrease of polymerase pausing in H1 deficient cells, which was confirmed by the generation of a metagene profile of RNApolII occupancy along the body of coding genes (Figure 20b): in H1-TKO cells, the RNApolII signal is slightly lower at the promoter and higher along the rest of the gene.

In spite of the lower calculated pausing index, we could not conclude that the initiation to elongation step was altered in H1-TKO cells. As seen in the meta-gene profile, the RNApolII signal is higher in H1-TKO cells not only inside the gene body, but also in the regions downstream of the promoter peak. Thus, the lower pausing index

could be just due to a better performance of the ChIP protocol in WT mES cells, which produced a higher signal-to-noise ratio.

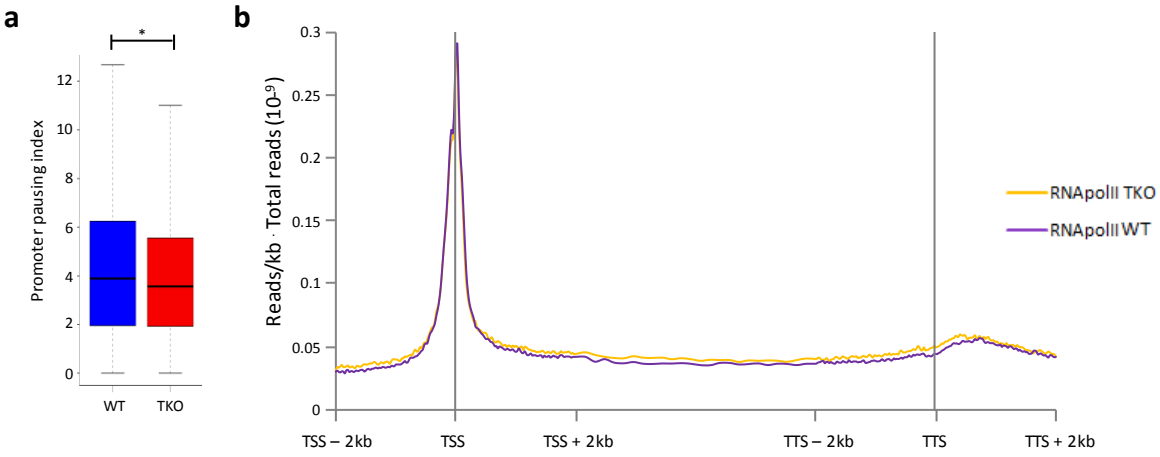


Figure 20: Analysis of RNA polymerase II promoter pausing **(a)** Boxplots showing the distribution of the pausing index of coding genes in WT and H1-TKO cells **(b)** Meta-gene profile of spike-in normalized RNApolII signal in a 4kb window around the TSS, along the gene body, and inside a 4kb window around the TTS of coding genes. *Mann-Whitney-Wilcoxon test p-value = $1.48 \cdot 10^{-13}$

3.3. Transcription elongation and termination defects

Transcription elongation can be perturbed by different motifs: RNApolII altered phosphorylation, R-loop formation, DNA supercoiling induction, or lack of certain RNA-binding proteins (Liang *et al.*, 2015; Tous and Aguilera, 2007; King *et al.*, 2013; Takeuchi *et al.*, 2018). Potentially, the reduced levels of histone H1 could be affecting any of these processes (we showed before that it triggers R-loop formation, Almeida *et al.*, 2018), so we searched for any evidence of elongation defects.

In some models of hindered transcriptional progression, like topoisomerase I inhibition or loss of the Sfpq RNA-binding protein, it has been described a long-gene transcriptopathy: RNApolII is unable to reach the transcription termination site of very long genes (>100 kb), so the overall amount of these transcripts drops (Takeuchi *et al.*, 2018; King *et al*., 2013). To check if a similar phenotype occurs in H1-TKO cells, we plotted cheRNA fold-change between WT and TKO conditions for all coding genes, classifying them by size (Figure 21a). We concluded that long genes tend to be less expressed when histone H1 levels are lower. This could be due to elongation problems, but could also reflect collisions of the transcriptional machinery with the

replication fork or other transcriptional complexes, that would increase with gene length.

Since transcription elongation seemed to be altered, we also investigated the consequences of the lack of histone H1 in transcription termination. Failures in the regulation of this process lead to transcriptional read-through, which is a common feature in several models of stress, cancer or viral infections (Vilborg *et al.*, 2017; Rutkowski *et al.*, 2015; Grosso *et al.*, 2015). In order to check if H1-TKO cells showed any defect in the regulation of termination, we measured the read-through index of the coding genes from the cheRNA-seq data. Read-through index is defined as the proportion of reads located 4 kb downstream the TTS and the number of reads located inside the gene body: this index is higher when transcription termination is defective. We plotted the index for the three replicates of WT and H1-TKO cells in two ways, as boxplots and as a cumulative plot (Figure 21b and 21c). In both cases, we could detect an increase of transcriptional read-through in the absence of the correct amounts of histone H1. However, the reduction on the levels of linker histone may not be the direct cause: the transcription termination defects could be the consequence of a general stress response triggered in these cells.

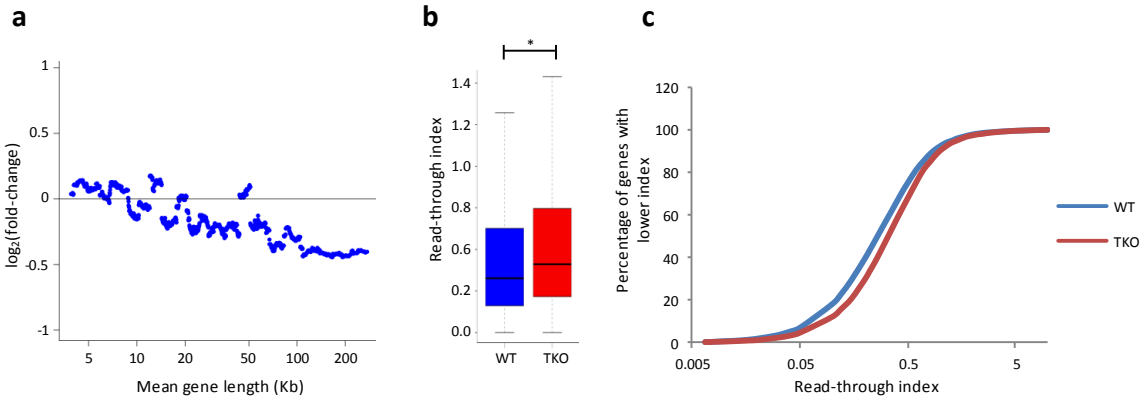


Figure 21: Analysis of transcription elongation and termination defects in H1-TKO cells **(a)** Plot showing the correlation between mean gene length (X-axis) and mean fold-change (Y-axis) in bins of 200 genes ordered by length **(b)** Boxplot showing the distribution of the log(read-through index) in WT and H1-TKO coding genes. *Mann-Whitney-Wilcoxon test p-value < $2.2 \cdot 10^{-16}$ **(c)** Cumulative plot representing, for different read-through index values, the percentage of genes with a lower index. A shift of the resulting curve to the right part of the plot corresponds to a general increase of the read-through index distribution.

4. Nucleosomal configuration in histone H1-TKO cells

4.1. Global nucleosome occupancy

The data so far supported a scenario where the pathological accumulation of non-coding transcripts and RNA polymerase II molecules in the chromatin of H1 deficient cells pose an obstacle to the passage of the replication fork, leading to replicative stress and DNA damage. However, the underlying molecular mechanism which connected the lack of linker histone with the accumulation of the transcripts was not clear. We considered two possible non-excluding explanations: a change in the promoter chromatin configuration, facilitating the recruitment of the transcription machinery at cryptic initiation sites; or a defect in the processing and degradation of the RNA, increasing the lifespan of normally unstable transcripts.

Regarding the first hypothesis, our previous analysis discarded substancial changes at the promoters of these genes in terms of histone post-translational modifications related to transcriptional regulation, including H3K4me1, H3K4me3 and H3K27me3 (Figures 12a and 15). Thus, we next decided to study the nucleosomal configuration and the potential differences in the accessibility of these regions, by sequencing preparations of MNase-digested chromatin.

Before looking for alterations in punctual sites of the genome, it was necessary to address if there was a global change in the nucleosomal occupancy between both cell types. This was achieved by performing three MNase replicates, adding two exogenous DNA fragments in known quantities as spike-in controls, and interrogating four representative genomic regions with high and low nucleosomal occupancy in mES by qPCR (see Materials and Methods, Table S6, Figure 22a). Next, the proportion between the amount of each genomic amplicon in a digested versus a not digested sample was calculated. Since these amounts are absolute quantifications (they have been normalized with a known concentration of the spike-in), the ratio is necessarily within the 0-1 range. These proportions were plotted for the two spike-in controls: Lma and Taq (Figure 22b).

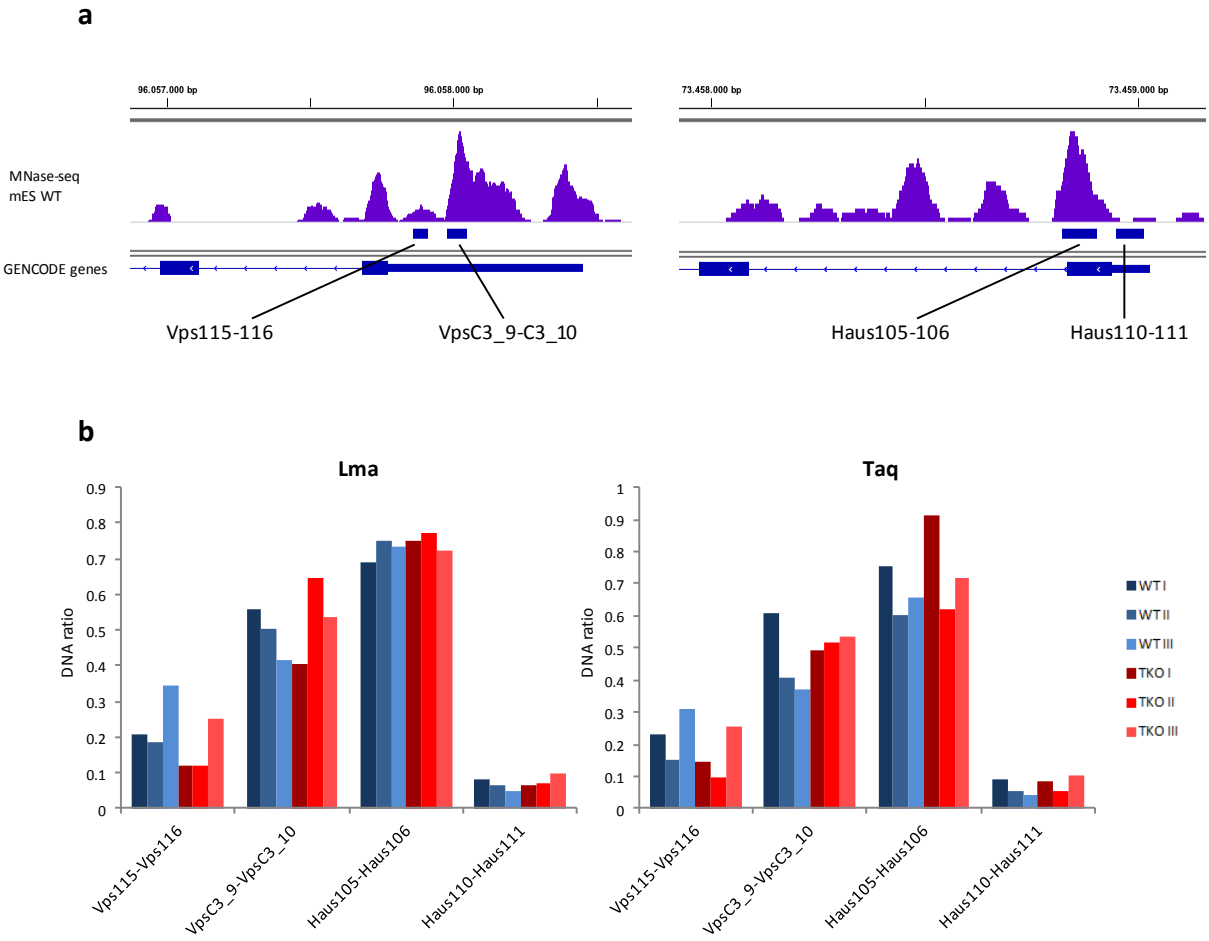


Figure 22: Analysis of global nucleosomal occupancy differences between WT and H1-TKO cells **(a)** Diagram of the four representative amplicons interrogated by qPCR. A mES WT MNase-seq track is included to estimate the expected digestion level for each region **(b)** Plots showing the ratios between the digested vs the undigested sample, for each one of the four amplicons tested in three biological replicates.

Since the difference in occupancy between WT and H1-TKO cells was not statistically significant (p-value > 0.05) for any of the four regions tested, with any of the two spike-in controls, we concluded that there was not a measurable change in the global nucleosome occupancy between the two conditions.

4.2. Differential nucleosomal fuzzyness

In order to study potential alterations in the nucleosomal landscape located in punctual sites of the genome, we performed the isolation and sequencing of MNase treated chromatin, coming from two replicates of WT and H1-TKO cells (Table S4). The computational software used to compare both conditions was DANPOS (see Materials and Methods), which first generates nucleosomal maps for each individual experiment, and then evaluates differences between them regarding three parameters: location shifts, positioning and occupancy.

The first parameter we analyzed was nucleosomal positioning. For each individual nucleosome in the genome, DANPOS calculates a positioning score, based on the shape and the width of the distribution of the reads, and compares it between two conditions. In the end, two lists of genomic coordinates were generated: one with the nucleosomes which are less positioned (or fuzzier) in H1-TKO cells, and another with those which are more positioned (less fuzzy).

Since histone H1 binding restrains nucleosome sliding (Pennings *et al.*, 1994), it could be predicted that reduced amounts of the histone cause a general decrease in nucleosomal positioning. To check this possibility, we used DANPOS to generate several lists of differentially positioned nucleosomes, setting different p-value cut-offs. For each list, the number of nucleosomes with increased fuzziness in WT or H1-TKO conditions was calculated and plotted (Figure 23a). As illustrated in the figure, there was indeed a decrease in the general positioning of nucleosomes in chromatin with reduced histone H1 levels: the stricter is the statistical threshold, the higher the proportion of fuzzier nucleosomes in H1-TKO cells.

Next, we studied where those differentially positioned nucleosomes were located in the genome. Setting a p-value threshold of 0.001, we obtained 55.608 differentially positioned nucleosomes. We quantified the number of them which matched promoters, gene-bodies and intergenic regions (defined from the cheRNA de novo assembled transcriptome), and compared them with the number of matches expected by chance (Figure 23b). Differentially positioned nucleosomes were statistically enriched within promoter regions. When different types of promoters were analyzed separately (Figure 23c), we found that all of them were enriched in differentially positioned nucleosomes.

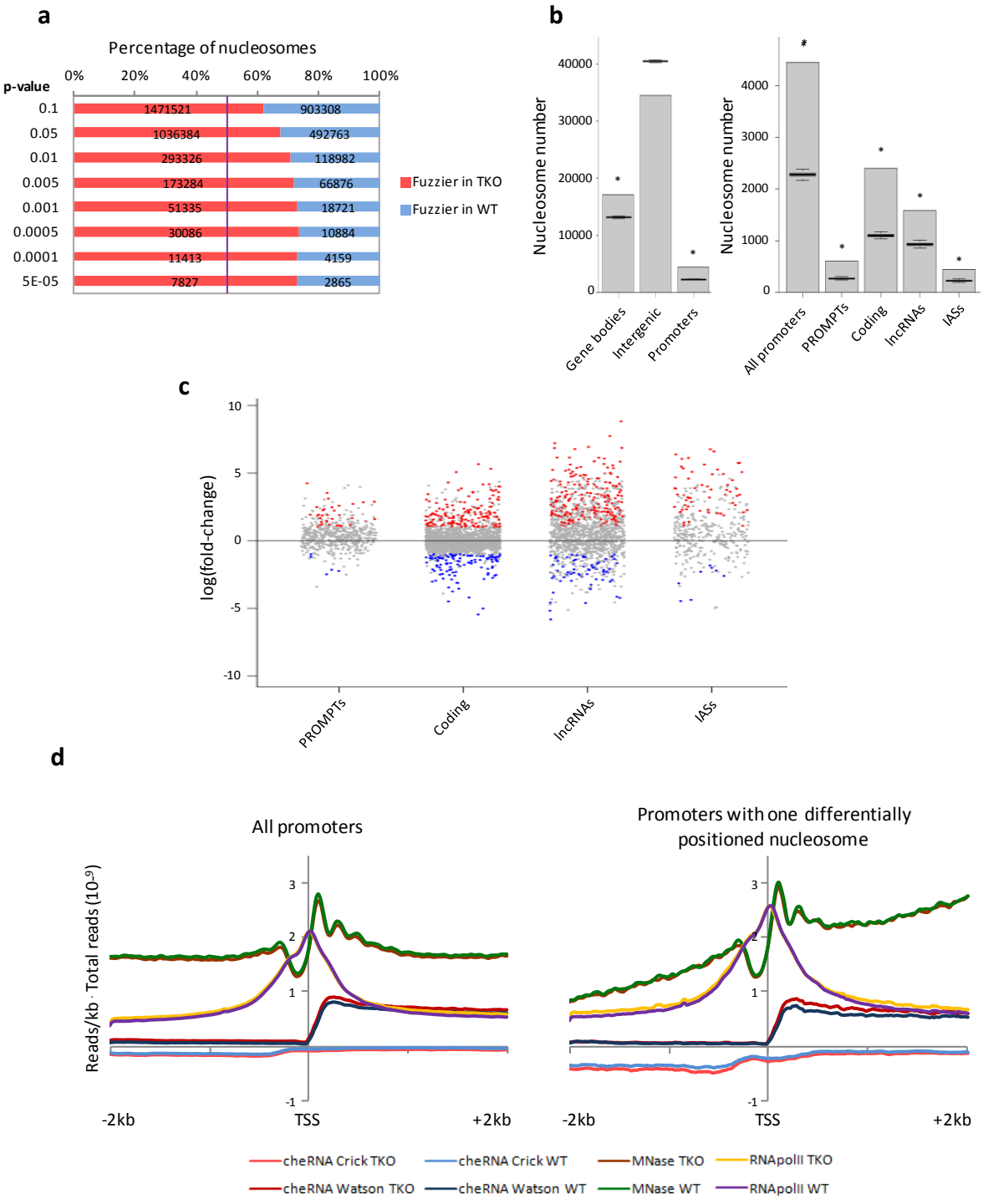


Figure 23: Characterization of nucleosomes with altered positioning **(a)** Plot showing the number of nucleosomes with increased or decreased positioning for different p-values **(b)** Number of differentially positioned nucleosomes in several genomic regions. Promoters are defined from the *de novo* assembled transcriptome, as the locus comprised from TSS-1kb to TSS+1kb; gene bodies encompass the region from TSS+1kb to TTS; the intergenic regions are the remainder genome. The boxplots represent the distribution obtained from 1000 random origin locations. *p-value<0.001 **(c)** Plot of the log(fold-change) of the transcripts associated to promoters with a differentially positioned nuclesome. The points in blue match the transcripts that are significantly downregulated; in red, the upregulated transcripts **(d)** Metaplots of MNase, RNApolII and cheRNA signals, plotted in 4kb windows around the TSS. RNApolII and MNase signals were multiplied by scale factors of 1:2 and 4, respectively, to facilitate the visualization in a single plot

Alterations in the nucleosomal configuration at promoters are normally coupled with a change in its transcriptional state. We therefore studied the promoters which contained at least one differentially positioned nucleosome, expecting to detect differences in the level of RNA generated from them. For each one of these promoters, we calculated and plotted the cheRNA fold-change of its associated gene, maintaining the classification in coding, lncRNAs, IASs and PROMPTs (Figure 23d). Surprisingly, no changes in their expression level could be detected, in spite of their altered chromatin configuration. This was further confirmed by plotting cheRNA and RNApolII signals in 4kb windows around the transcription start sites of these genes, which showed no differences between WT and H1 deficient cells (Figure 23e).

4.3. Differential nucleosomal occupancy

Another parameter that DANPOS measures is nucleosomal occupancy, by comparing the number of reads within each individual nucleosome. In this case, with a p-value cut-off of 0.001, we found 20.323 regions with differential occupancy. To study their genomic location, we repeated the previous analysis, evaluating their enrichment at promoters, gene bodies and intergenic regions (Figure 24a). We detected that differentially occupied nucleosomes were preferentially located at promoters, in a similar fashion than differentially positioned nucleosomes. This enrichment was also independent on the type of promoter: it occurs around coding, lncRNAs, PROMPTs and IASs transcription start sites (Figure 24b).

Since we found no differences in global nucleosome occupancies in H1-TKO cells (Figure 22), it could be predicted that the alterations in specific sites of the genome would include both increases and decreases in occupancy. This was confirmed analyzing the list of differentially occupied nucleosomes: 48% of them show a decrease, and 52% an increase in occupancy. However, these percentages were drastically changed when taking into account only the differential nucleosomes located at promoters (Figure 24c). Almost all of them have a lower occupancy in H1-TKO cells, what suggests that these promoters have a more open chromatin configuration.

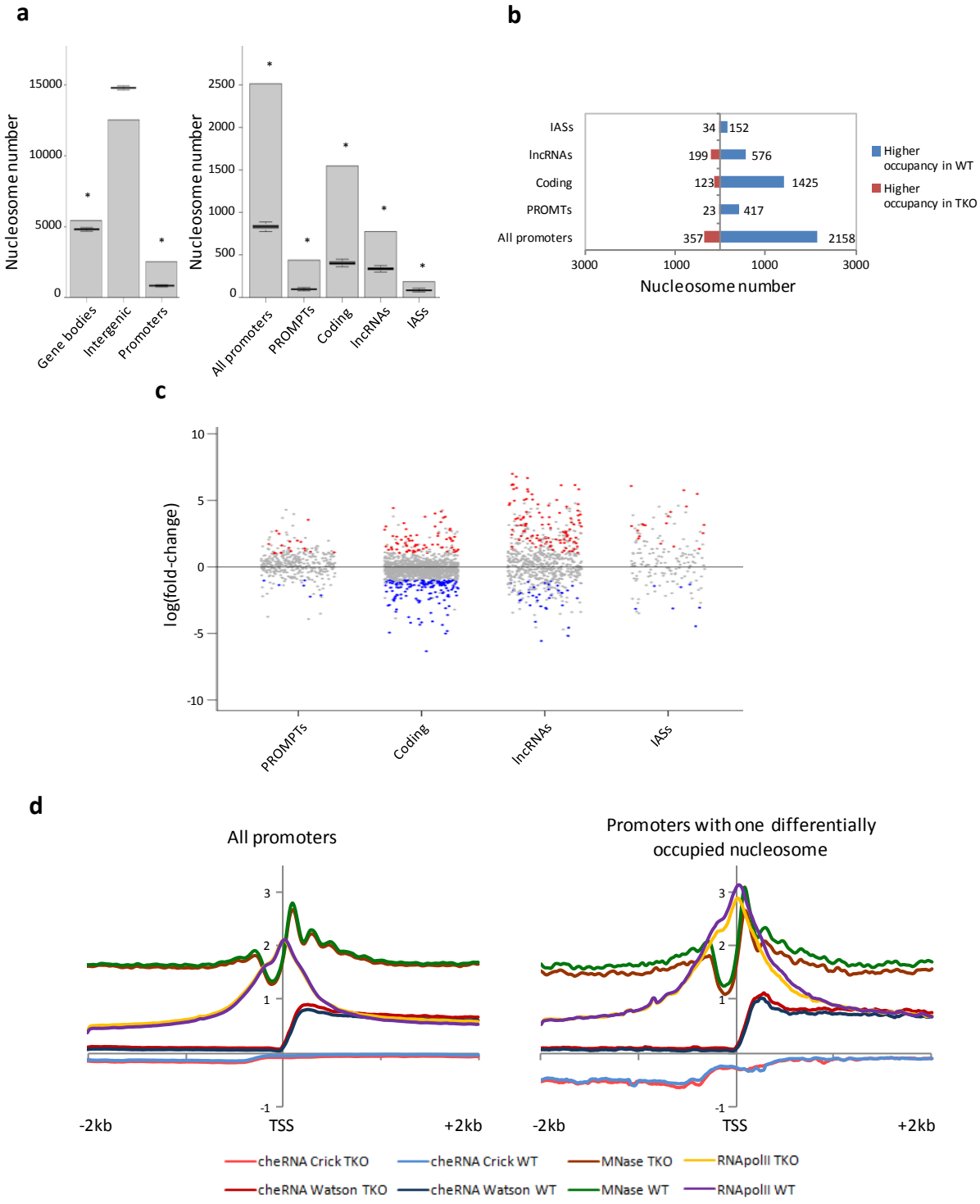


Figure 24: Characterization of nucleosomes with altered occupancy **(a)** Number of differentially positioned nucleosomes in several genomic regions. Promoters are defined from the de novo assembled transcriptome, as the locus comprised from TSS-1kb to TSS+1kb; gene bodies encompass the region from TSS+1kb to TTS; the intergenic regions are the remainder genome. The boxplots represent the distribution obtained from 1000 random origin locations. *p-value<0.001 **(b)** Number of promoters with increased or decreased nucleosomal occupancy **(c)** Plots representing the log(fold-change) of the transcripts associated to promoters with a differentially occupied nucleosome. The points in blue match the transcripts that are significantly downregulated; in red, the upregulated transcripts **(d)** Metaplots of MNase, RNApolII and cheRNA signals, plotted in 4kb windows around the TSS. RNApolII and MNase signals were multiplied by scale factors of 1:2 and 4, respectively, to facilitate the visualization in a single plot

Reduced nucleosome occupancy at promoters has been described to increase its accessibility to DNA binding proteins, resulting in a more efficient recruitment of the transcription machinery and a higher transcription rate (reviewed in Bai and Morozov, 2010). Thus, we could expect that this group of promoters gave rise to the differentially expressed transcripts we had found in the chromatin enriched RNA analysis. To check this, we plotted the fold-change of the transcripts associated to the promoters with reduced nucleosomal occupancy (Figure 24d). As in the previous analysis on differentially positioned nucleosomes, we did not detect any significant change in the levels of these transcripts between WT and H1-TKO cells. Again, we confirmed these observations by plotting cheRNA and RNApolII signals in 4kb windows around their TSS (Figure 24e).

5. Phenotypic similarities between H1-TKO and *Exosc3* knock-down cells

5.1. Differential RNA-seq analysis in Exosc3 knock-down cells

Since we didn't find any correlation between differentially accumulated transcripts in H1-TKO cells chromatin and the epigenetic state or nucleosomal configuration of their TSS-surrounding regions, we next studied if there was any alteration in their turnover. The categories of non-coding transcripts that were perturbed upon the reduction of histone H1 levels are known targets of the RNA surveillance complex, the RNA exosome (Flynn *et al.*, 2011; Pefanis *et al.*, 2015; Schlackow *et al.*, 2017). Moreover, some of the transcriptional phenotypes that are present in H1-TKO cells, like elongation defects and transcriptional read-through (Figure 21a, 21b and 21c), are also distinctive characteristics of several mutants of exosome components (Luna *et al.*, 2005; Lemay *et al.*, 2014). All these observations suggested a possible connection between the histone H1 and the RNA exosome complex, which could be necessary to degrade certain categories of transcripts.

As a first approach to check this hypothesis, we reanalyzed published RNA-seq data derived from mES cells expressing a shRNA to silence *Exosc3* (Pefanis *et al.*, 2015), a core component of the RNA exosome. We aligned the total RNA-seq reads from the shExosc3 cells and their WT counterpart to the *de novo* assembled transcriptome that we had obtained from the cheRNA-seq experiments and performed

a differential expression analysis. In spite of the different RNA populations that had been isolated, we detected a similar pattern of non-coding transcript accumulation to the one we had described for H1-TKO cells (Figure 25a and 25b).

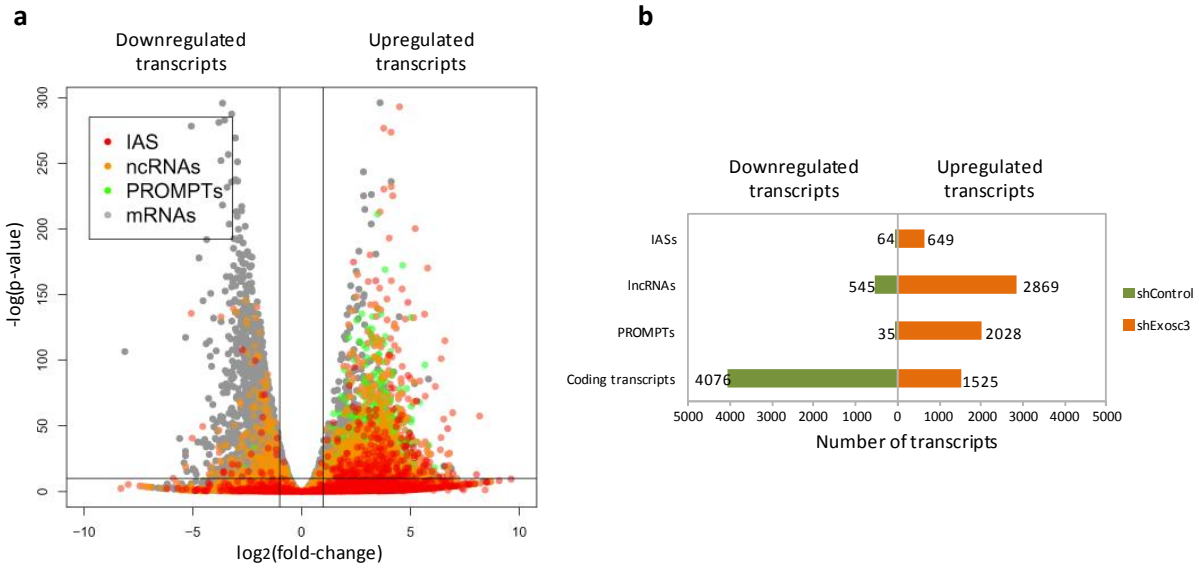


Figure 25: Differential expression analysis in shControl and shExosc3 cells **(a)** Vulcano plot showing the log(fold-change) and the -log(p-value) for each transcript in Exosc3 knock-down cells **(b)** Number of differentially expressed transcripts between shControl and shExosc3 cells for each category

5.2. Replication defects in Exosc3 knock-down cells

To further investigate the possible connection between histone H1 and the nuclear exosome complex, we tested whether the DNA replication alterations we had found in H1-TKO cells were recapitulated in exosome-downregulated cells. To this aim, we generated a mES cell line with a stable integration of a shRNA that targets *Exosc3* mRNA levels in WT mES cells, and checked the reduction in EXOSC3 levels by Western Blot (Figure 26a). Next, we performed fiber stretching analysis of these cells, finding a remarkably similar phenotype to that caused by the reduction in the levels of histone H1 (Figure 4). In shExosc3 cells, the velocity of the replication forks decreases in a transcription dependent manner: the inhibition of RNA polymerase II after a 3h treatment with DRB restores the fork rate measure at the WT levels (Figure 26b, untreated and DRB columns, Table S5). In an analogous way to H1-TKO cells, the re-start of the transcriptional activity one hour after the removal of DRB triggers again the low fork rate phenotype (Figure 26b, DRB-release columns, table S5).

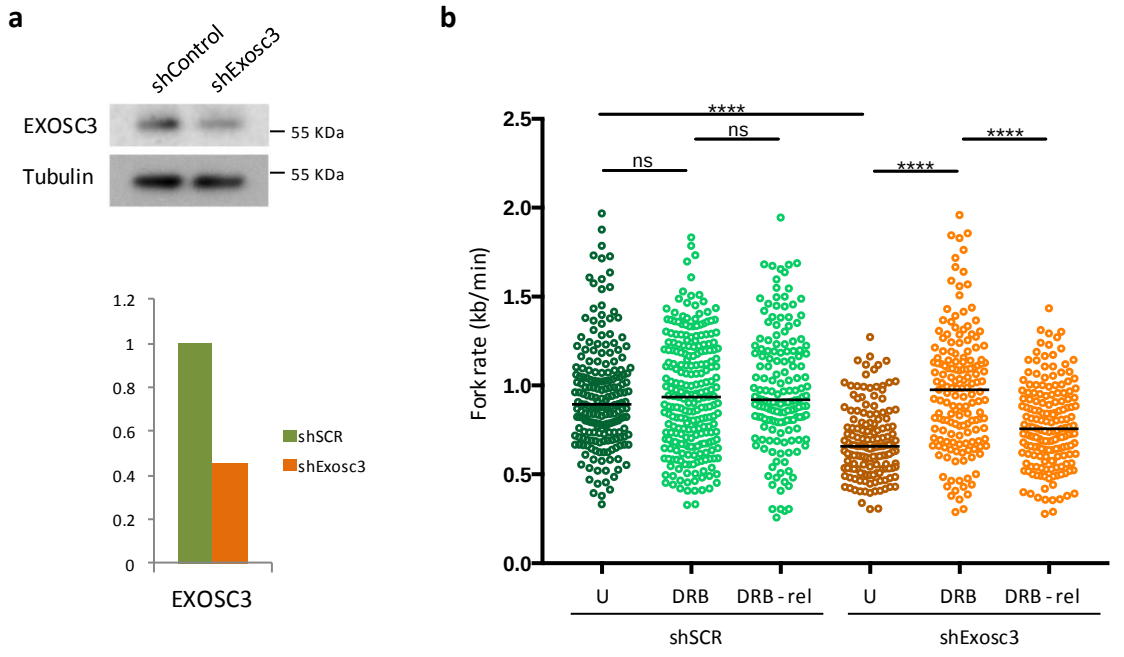


Figure 26: Analysis of DNA replication in Exosc3 knock-down cells **(a)** Western blot showing the downregulation of the levels of EXOSC3 protein (upper panel) and the corresponding quantification (lower panel) **(b)** Measure of replication fork rate in shControl and shExosc3 cells, untreated (U), treated with DRB for 3 hours (DRB), and 1 hour after DRB block release (DRB-rel). Medians are represented. Statistical differences between distributions were assessed with the Mann–Whitney rank sum test. ****p-value<0.0001.

DISCUSSION

1. Characterization of differentially expressed transcripts in H1-TKO cells

In previous studies in our group, we unveiled a connection between the lack of the correct amounts of histone H1 and the generation of wide-spread transcription-replication conflicts. Since transcriptional studies in these cells had shown very limited changes regarding mRNA expression (Fan *et al.*, 2005; Geeven *et al.*, 2015), we decided to investigate whether these conflicts were due to a change in the levels of unstable non-coding RNAs, which are lost during standard RNA obtention procedures. In order to specifically isolate these kind of RNAs, we performed a cellular fractionation to obtain chromatin enriched RNA. cheRNAs can be bound to chromatin through different mechanisms, but more than 75% of them are nascent RNAs tethered to the DNA by an RNA polymerase II molecule (Werner and Ruthenburg, 2015). The most remarkable finding was the widespread accumulation of several types of non-coding transcripts in the chromatin of H1-TKO cells, including lncRNAs, internal antisense transcripts and PROMPTs (Figures 9b and 9c), but differences regarding coding transcripts levels were also detected.

To confirm that the differential transcripts were actually nascent RNA, we performed a RNA polymerase II ChIP. We could detect that any increase in the amount of all types of transcripts was correlated with higher levels of RNA polymerase II recruitment (Figure 19), demonstrating that RNApolII was the protein that anchors these transcripts to chromatin. Moreover, the accumulation of non-coding transcripts was coupled to the presence of RNA polymerase II molecules along intergenic regions (Figures 18b and 18c), possibly triggering the conflicts with the replication machinery.

As stated before, previous transcriptomic analysis in these cells had found very few alterations in coding RNAs: 29 genes were identified as differentially expressed by microarray hybridization in Fan *et al.*, 2005 and 599 by polyadenylated RNA-seq in Geeven *et al.*, 2015, while by cheRNA analysis we detected 2904 differential coding transcripts. It is difficult to explain this discrepancy, but it could be related to the different RNA population that is being studied: cheRNAs have to go through additional regulation processes after being released from chromatin, including the decision between the alternative pathways of RNA decay or nucleo-cytoplasmic export (Tudek *et al.*, 2018).

When we analyzed the features of the differentially accumulated coding transcripts, we found that they were involved in differentiation and development (Figure 11a), and their promoters tend to be enriched in Polycomb features: H3K27me3 and Suz12 binding sites (Figures 11b and 11c). This agrees well with the previous finding of a connection between histone H1 and the Polycomb complex: histone H1 binds nucleosomes containing trimethylated H3K27, leading to the transcriptional silencing of Polycomb-regulated promoters (Kim *et al.*, 2015). However, these alterations could also be due to an indirect effect of the lack of histone H1: H1-TKO cells show an altered differentiation potential (Zhang *et al.*, 2012), what could indirectly affect the expression of genes involved in development, mainly regulated by the Polycomb complex (reviewed in Aloia *et al.*, 2013).

Beyond development genes, another set of differentially accumulated transcripts were integrated by antiviral defense and interferon-response elements (Figure 11a). In a previous transcriptomic analysis in T47D H1-knockdown cells (Izquierdo-Boulstridge *et al.*, 2017), it had been reported that H1 is essential for heterochromatic repeats and endogenous retroviruses silencing. When the amount of histone H1 was reduced, double-stranded RNA was synthesized from these repeats and exported to the cytoplasm, triggering an interferon response and the consequent severe impairment of cell growth. In H1-TKO cells, the derepression of heterochromatic repeats was also described (Cao *et al.*, 2013); however, no interferon-mediated apoptosis or growth defects can be detected. This can be due to the fact that, although we have found interferon related genes whose expression is upregulated, all of them were dsRNA sensors (*Oas1a*, *Oas1c*, *Oas1g* and *Oasl2*): we did not detect overexpression of the genes that are responsible of signal transduction or apoptosis induction. Accordingly, the absence of any phenotypic effect in response to dsRNA accumulation in the cytoplasm of H1-TKO cells is not surprising, since embryonic stem cells are unable to trigger a complete interferon response under any circumstances, including viral infections (Wang *et al.*, 2013) or dsRNA analogs transfection (Witteveldt *et al.*, 2019).

Regarding non-coding transcripts, their characterization is more challenging, since most of them have been *de novo* annotated and their function is unknown. In previous studies, non-coding cheRNAs had been found to work as enhancer RNAs, colocalizing with nearby promoters and increasing their activity (Werner *et al.*, 2017).

To test if the lncRNAs we had annotated were actually eRNAs, we measured the activity of the proximal promoters, which was increased compared to promoters located far away from a lncRNA (Figure 13a). Moreover, their TSS were preferentially located in regions with high H3K4 monomethylation levels (Figures 13b and 15), the most commonly used epigenetic feature to predict enhancer function (reviewed in Shlyueva *et al.*, 2014).

We tried to get a more precise estimation of how many of the non-coding transcripts we had assembled were eRNAs, but we couldn't find a reliable way of computationally discriminating between enhancers and non-coding promoters. The first approach consisted in measuring the H3K4me1/H3K4me3 ratio, which has been described to be higher in enhancers (Robertson *et al.*, 2008). However, there are enhancers with H3K4me3 levels comparable to those in promoters (Pekowska *et al.*, 2011), and H3K4me1 is also involved in the silencing of inactive coding promoters (Cheng *et al.*, 2014). Another epigenetic mark to predict enhancers is H3K27ac, which has been traditionally regarded as a mark of active enhancers (Creyghton *et al.*, 2010), but it is also present at promoters (Wang *et al.*, 2008). In addition to epigenetic modifications, the presence of certain DNA binding proteins is used as mark for enhancer function, like p300 or CHD7 (Visel *et al.*, 2009), but both of them are also found around the TSS of annotated promoters (Vo and Goodman, 2001; Schnetz *et al.*, 2010). Other enhancer features, like nucleosome depleted regions or DNaseI hypersensitivity, were actually described first at promoters. To overcome these limitations in computationally differentiation between enhancers and non-coding promoters, we decided to look for a database annotation of enhancers, and check if our assembled lncRNAs were synthesized from them. In the case of the dbSUPER mouse embryonic stem cells superenhancer annotation (Khan and Zhang, 2016), we could detect that 100% of the enhancers matched the promoter of at least one *de novo* assembled lncRNA.

The presence of chromatin enriched eRNAs in all the annotated enhancers, together with the impossibility of finding a reliable mark to computationally differentiate them from promoters, seem to agree with the hypothesis that enhancer and promoters are actually the same type of regulatory regions, and that there are no significant differences between them regarding their cellular function (Andersson *et al.*, 2015). This theory is based on the capacity of enhancers to recruit RNA polymerase II and

initiate transcription (De Santa *et al.*, 2010; Kim *et al.*, 2010; Arner *et al.*, 2015), but also in the ability of promoters to increase the transcriptional activity of nearby elements: in several functional assays, a large fraction of the genomic regions with enhancer activity were actually coding promoters (Zabidi *et al.*, 2015; Barakat *et al.*, 2018). In this context, discriminating eRNAs within our group of assembled lncRNAs would not make sense, since the only differences between them would be their stability and function, not the features of their promoter region.

Recapitulating, we can only assert that a significant fraction of the newly assembled non-coding transcripts are enhancer RNAs, but no better estimation of the real percentage could be calculated. In any case, both eRNAs (Melo *et al.*, 2013; Li *et al.*, 2013) and other types of non-coding transcripts (Wang *et al.*, 2011; Lai *et al.*, 2013) have been shown to increase the transcriptional activity of proximal coding promoters by different cis-regulatory mechanisms. Thus, we decided to study the function of coding genes located next to a differentially accumulated lncRNA in H1-TKO cells, which were involved in differentiation and transcription from RNApolII promoters (Figure 14a). This last group included transcription factors also related to differentiation. Next, we wondered if the change of the lncRNA level had any consequence on the activity of the proximal promoter. For downregulated lncRNAs, we found the expected phenotype: the nearby promoters were also downregulated. Surprisingly, coding promoters located near an upregulated lncRNA showed almost no changes in their expression (Figure 14b). The fact that the accumulation of lncRNAs in the chromatin of H1-TKO cells is not triggering the physiological transcriptional response suggests that this accumulation is a pathological consequence of the lack of correct amounts of histone H1.

One of the most notable phenotype of H1-TKO cells is their inability to differentiate, due to the role of histone H1 in the repression of several pluripotency genes, like *Nanog* and *Oct4* (Zhang *et al.*, 2012). In our cheRNA analysis, we found no differences regarding the expression of *Nanog*, *Oct4* and *Sox2* coding genes between WT and H1-TKO cells; however, all of them had an upregulated lncRNA rising from their respective enhancers (Figure 14c). An interesting possibility would be that the presence of this eRNAs constitutes an obstacle for the shutdown of these genes: maybe by knocking-down these three eRNAs the differentiation phenotype of these cells could be rescued.

2. Role of histone H1 in transcriptional regulation

The simplest explanation for the accumulation of transcripts in the chromatin of H1-TKO cells involved a change in the epigenetic state of their promoters caused by the lack of histone H1. As mentioned before, transcriptional activation of a promoter requires the eviction of H1 from it (see introduction, section 4). However, few models which mechanistically link the absence of histone H1 with the subsequent activation have been proposed. Two of them are the H1-dependent inhibition of the H3K4 methylase SET7/9, and the recruitment of DNMT1 and DNMT3B (Yang *et al.*, 2013).

DNMT1 and DNMT3B are DNA methylases which are recruited to chromatin by histone H1. DNA methylation is a modification of the cytosines of the DNA present in certain regulatory regions, favoring a closed chromatin structure that diminishes the accessibility of the transcription machinery. DNA methylation has been studied in H1-TKO cells (Geeven *et al.*, 2015), finding a great number of regions with differential methylation patterns: 49% of coding and 25% of total non-coding promoters contain a differentially methylated region (DMR). When we analyzed specifically the promoters of differentially expressed transcripts in H1-TKO cells, these percentages were completely similar: 46% of the coding and 27% of the non-coding transcripts contain a DMR in their promoter. This lack of correspondence between DNA methylation and transcriptional alterations in H1 deficient cells was also observed in T47D H1 knock-downs (Izquierdo-Boulstridge *et al.*, 2017).

SET7/9 is a histone methyltransferase involved in the methylation of H3K4, a epigenetic mark related to transcriptional activation. The binding of SET7/9 to chromatin is inhibited by the presence of histone H1 (Yang *et al.*, 2013), so the transcriptional alterations in H1 deficient cells could be explained by the hypermethylation of H3K4, caused by the overactivation of the methyltransferase in regulatory regions of the genome. To check this hypothesis, we made use of H3K4me1 and H3K4me3 ChIP-seq published data, obtained from both WT and H1-TKO cells (Geeven *et al.*, 2015), assessing if they were enriched in the promoters of differentially expressed transcripts. In agreement with the original study, we found no correlation between H3K4 methylation and transcriptional alterations in these cells (Figures 12a and 15).

Another possible mechanism for H1 direct regulation of RNA polymerase recruitment has been described for p53 targets. In the promoters of silenced target genes, histone H1 form a complex with p53, avoiding the action of the histone acetyl-transferase p300 (Kim *et al.*, 2008). Since many differentiation-related genes are targets of p53, the differentially expressed transcripts in H1 deficient mES cells (also involved in differentiation and development) could be regulated in this way. However, this p300-dependent regulation mechanism also involves an increase in H3K4 methylation (Tang *et al.*, 2013), which is not present at the promoters of these transcripts.

Besides transcription initiation, histone H1 could also play a role in the promoter pausing release of the RNA polymerase II. The variant H1c interacts with the Ser2 phosphorylated form of the RNApolII CTD, inducing the recruitment of the CUL4A E3 ubiquitin ligase and the PAF1C elongation complex, and displaying a positive effect on the expression of certain genes (Kim *et al.*, 2013). Since PAF1C stimulates the pause release of the polymerase (Chen *et al.*, 2015), it would be possible that the transcriptional defects of H1 deficient cells were due to the reduced recruitment of PAF1C during the initiation to elongation step. However, several evidences argue against this possibility. In the first place, we found almost no difference in promoter pausing in the RNApolII ChIP-seq data (Figure 20); at most, it could be detected a slight decrease in the pausing index, when a reduced binding of PAF1C actually increases it (Chen *et al.*, 2015; Shivji *et al.*, 2018). Moreover, the phenotype caused by the lack of the protein PAF1 (key component of the PAF1 complex) is the opposite of that described for H1-TKO cells: in yeast, the levels of unstable non-coding transcripts decreased upon deletion of *PAF1* gene (Ellison *et al.*, 2019). Conversely, the reduced recruitment of PAF1C could possibly explain the elongation defects, since it is needed to maintain an appropriate elongation rate in mammalian cells (Hou *et al.*, 2019).

Consequently, none of the described models for histone H1 transcriptional regulation seemed to explain the accumulation of certain types of transcripts in the chromatin in H1-TKO cells. Since we couldn't detect any epigenetic alteration in this set of promoters, we decided to check if the lack of linker histone was increasing the accessibility of the chromatin in these regions, facilitating the recruitment of basal transcription factors. In order to do that, we studied the nucleosomal configuration of H1-TKO cells. The influence of nucleosome positioning and occupancy on

transcriptional regulation has been widely characterized (reviewed in Bai and Morozov, 2010). In general, transcriptional activation requires a remodeling of the chromatin, involving the displacement or removal of nucleosomes to make the TSS accessible for the RNA polymerase II binding. Histone H1 has been described to stabilize the nucleosome binding to DNA (Harshman *et al.*, 2013), hence increasing nucleosome occupancy in regulatory regions (Hu *et al.*, 2018). A possible link between the lack of histone H1 and the overexpression of certain transcripts could be simply the increase in accessibility in a specific set of sensitive promoters.

To test this hypothesis, we generated and sequenced MNase digestions of the chromatin of WT and H1-TKO cells. The software DANPOS allowed us to map nucleosomes with altered positioning and occupancy. In agreement with previous studies (West *et al.*, 2014; Chen *et al.*, 2013), the altered nucleosomes were enriched at the promoters of all types of transcripts (Figures 23b and 24a). The most notable difference between the WT and the H1-TKO condition regarding nucleosome configuration was a clear decrease in the nucleosomal occupancy in a group of promoters of H1 deficient cells (Figure 24c). As stated before, this decreased occupancy was expected to activate the transcription of these genes; however, we could see that neither the alterations of positioning nor the drop in occupancy had any effect on the transcriptional output of H1-TKO cells (Figures 23c, 23d, 24c and 24d).

One possible explanation for this could be that, arguably, MNase-seq is not the best technique to study the accessibility of the chromatin. In recent studies (Mieczkowski *et al.*, 2016; Mueller *et al.*, 2017), it has been demonstrated that there can be changes in accessibility in certain promoters which are not reflected by the MNase-seq calculated occupancy. The authors claim that an Assay for Transposase Accessible Chromatin (ATAC-seq) constitutes a better measure, since changes in ATAC-seq signal are better correlated with transcriptional activity. Thus, it could be possible that the differentially expressed transcripts in H1-TKO cells would actually show alterations in an ATAC-seq experiment. Nevertheless, this was already performed in T47D histone H1 knockdowns (Izquierdo-Boulstridge *et al.*, 2017), finding just a slight tendency of accessibility peaks to be enriched inside differentially expressed promoters (below statistical significance thresholds).

Taking all these results into account, we have not found any epigenetic feature that is significantly altered at the promoters of the transcripts that are differentially

accumulated in the chromatin of H1-TKO cells. We can't still discard that the lack of histone H1 has any direct consequence on the activity of certain promoters, but it seems clear that at least another different mechanism is needed to explain the upregulation of non-coding transcripts in these cells. As mentioned previously, we decided to investigate if the decrease of histone H1 amounts could be altering the degradation of these transcripts, instead of their synthesis, increasing their mean residence time in chromatin.

3. Role of histone H1 in RNA metabolism and degradation

The degradation of intrinsically unstable RNAs inside the nucleus of the cell is mainly performed by the nuclear exosome complex. Among the targets of this complex are several types of cryptic transcripts: PROMPTs (Flynn *et al.*, 2011), enhancer RNAs (Pefanis *et al.*, 2015), polyadenilated lncRNAs (Schlackow *et al.*, 2017) and transcripts generated from heterochromatic repetitive elements (Wang *et al.*, 2008). Since the type of transcripts that are upregulated in the chromatin of H1-TKO cells are normally exosome targets, we decided to study if their degradation mechanism was altered in these cells. As a first approach, we made use of total RNA-seq published data coming from *Exosc3* knock-down embrionary stem cells (Pefanis *et al.*, 2015). When we aligned the RNA-seq reads to our cheRNA *de novo* assembled transcriptome, we could see that the pattern of non-coding transcriptional alteration was similar between the knock-down of the exosome complex and the deletion of the histone H1 variants (Figure 25). Thus, an interesting possibility would be that either histone H1 directly, or the correct amounts of histone H1, are needed for the correct recognition of these transcripts by the RNA degradation machinery.

Besides unstable non-coding transcripts, several coding RNAs have been discovered to be targets of the exosome complex. During stem cell differentiation, it has been shown that the regulation of the transcriptional change of mRNAs is a combination of the alteration of their synthesis rate and the regulation of their decay (Lloret-Llinares *et al.*, 2018). The mRNAs that are preferentially regulated at the stability level are exosome targets, and share only one feature: they are lowly expressed transcripts in normal conditions. This resembles the situation found in H1-TKO cells, where the up-regulated coding transcripts have low expression levels in WT

mES cells (Figure 12b) and are related to differentiation and development (Figure 11a), so histone H1 could also play a role in the degradation of these coding transcripts by the exosome complex.

To further investigate a possible link between histone H1 and the exosome complex, we generated a *Exosc3* knock-down mES cell line in the WT genetic background of our cells. We found that at least some of the replicative defects present in *Exosc3*-KD cells were reminiscent to those of H1-TKO cells: we could detect that the fork rate decreases upon the reduction of the nuclear exosome activity (Figure 26b). This alteration was dependent on the presence of active transcription: in both cases, it was reverted when DRB was added to the culture medium. Inter-origin distance and fork asymmetry in *Exosc3*-KD cells will be also studied in future experiments.

The replicative phenotype of H1-TKO cells is not the only one that resembles the situation in knock-downs of the RNA exosome complex. In every transcription event, there is a certain probability of failure of the standard termination process at the poly-adenylation site. In these cases, there are fail-safe mechanisms that ensure the disassembly of the transcriptional machinery, all of them linked to the exosome activity (Colin *et al.*, 2014; Rondón *et al.*, 2009; Gudipati *et al.*, 2012; Webb *et al.*, 2014). In some cases, the switching between these two modes of termination can be a way of regulating the expression of the transcript. When the exosome activity is compromised due to the knock-down of the dis3 gene in *Schizosaccharomyces pombe* cells, the fail-safe mechanism is not active, leading to transcriptional read-through and the accumulation of 3'-extended RNAs (Lemay *et al.*, 2014). Transcriptional read-though has also been detected in mouse B cells upon the knock-down of the *Exosc3* gene (Pefanis *et al.*, 2014). We found 3'-extended RNAs when histone H1 amounts are reduced (Figures 21b and 21c), further strengthening the possible influence of H1 in the RNA exosome activity; however, transcriptional read-through is present in different models of cellular stress, like cancer (Grosso *et al.*, 2015) or viral infections (Rutkowski *et al.*, 2015), so it could also be due to a general stress response triggered in these cells.

In addition to termination defects, we have detected alterations in transcriptional elongation when histone H1 levels are reduced (Figure 21a). Long genes are less expressed than short genes, probably due to the inability of RNA polymerase II to reach the canonical transcription termination site, consequently disassembling from

chromatin inside the gene body. A similar phenomenon has been described in *S. cerevisiae rrp6Δ* mutants (Luna *et al.*, 2005). In this study, the percentage of RNA polymerase II that reached the 3' end of a reporter gene was measured: when RRP6 gene was deleted, this percentage dropped from near 100% in WT conditions to 65%–76%. However, we can't directly assign the effect in elongation to an interaction between histone H1 and exosome activity: other reasons could explain early RNA polymerase II disassembly, like collisions with other RNApolII molecules or replication forks, or the direct interaction of histone H1 with the FACT (facilitates chromatin transcription) complex (Zhang *et al.*, 2016), which facilitates nucleosome removal of the template during transcription elongation (Hsieh *et al.*, 2013).

Another shared aspect between H1-TKO cells and models with deficient exosome activity is the de-repression of heterochromatin repeats. As mentioned earlier, the lack of histone H1 causes an upregulation of transcripts coming from centromeric repeats (Cao *et al.*, 2013), and the exosome complex also plays a role in the degradation of this kind of transcripts (Bühler, 2009).

To sum up, histone H1 deficient mES cells show a very similar phenotype to exosome mutant models regarding DNA replication defects, non-coding transcript accumulation, mRNA transcriptional regulation, transcriptional read-through, transcription elongation and centromeric silencing failure. Still, we have no direct evidence of a decreased decay rate for the transcripts that are accumulated in the chromatin of H1-TKO cells. To finally demonstrate that histone H1 is important for the regulation of the metabolism of certain transcripts, we are planning to measure the time that is required for the degradation of some of the differentially accumulated transcripts after RNA polymerase II inhibition, comparing this time between H1-TKO and WT conditions.

In spite of the phenotypic similarity, the molecular mechanism that explains the possible connection between the linker histone and the exosome complex remains unknown. One possibility would be that histone H1 physically interacts with some proteins involved in the RNA targeting to any of the exosome components. In agreement with this hypothesis, $H1^0$ variant has been shown to interact directly with numerous RNA binding proteins which could potentially regulate RNA stability and decay (Kalashnikova *et al.*, 2013; Szerlong *et al.*, 2015), including splicing factors, hnRNPs, poly-A binding proteins and RNA helicases. Most of these interactions are

related to the role of $H1^0$ in ribosomal RNA synthesis and processing at the nucleolus (Kowalski 2016), but some of these proteins bind specifically mRNAs, such as PABPC1, PRPF40a and Matrin-3. Unfortunately, little is known about the protein interactome of the three deleted variants in our mES cells (H1c, H1d and H1e). Anyway, several RBPs have been found to form a complex with variant H1c. Among them are TPR, a component of the nuclear pore complex involved in RNA export, and the poly-A binding proteins PABPC3 and PABPC4, which regulate the stability of mRNAs in the cytoplasm (Zhang *et al.*, 2016).

4. RNA processing and genomic instability

H1-TKO cells show several DNA replication defects due to transcription-replication conflicts, as they are reverted upon inhibition of RNA polymerase II (Figure 4). In these cells, we have underscored an accumulation of different types of non-coding transcripts in chromatin, which could be responsible of the conflicts with the replication machinery (Figure 9b and 9c). For decades, transcription has been regarded as a source of mutagenesis and instability, i.e., high levels of transcription have been linked to an increase in spontaneous mutations rates (Datta and Jinks-Robertson, 1995). However, it does not seem probable that the increase of non-coding transcription rate would cause any problem to the cell homeostasis: some coding genes are still several orders of magnitude more transcribed than the set of differentially accumulated transcripts. Since we have found that the knock-down of the exosome component *Exosc3* leads to a partially comparable replicative phenotype to that present in H1-TKO cells (Figure 26b), an interesting possibility could be that the cause of the conflicts is a delay in the processing and decay of certain RNAs, what would also explain the non-coding transcript accumulation.

There are multiple studies that demonstrate a molecular link between defects in RNA processing and genomic instability, including alterations in RNA cleavage and polyadenylation (Stirling *et al.*, 2012), transcription termination (Becherel *et al.*, 2013; Morales *et al.*, 2016), splicing (Li and Manley, 2005; Chakraborty *et al.*, 2018), or nuclear export (Hodroj *et al.*, 2017; García-Benítez *et al.*, 2017; Teloni *et al.*, 2019). In all these cases, the instability is mediated by the formation of R-loops, which trigger ssDNA exposure, increased mutagenesis, hyper-recombination and conflicts with the replication machinery (Crossley *et al.*, 2019). Importantly, some models of exosome

compromised activity also showed this R-loop mediated genome instability and hyper-recombination, like the yeast knock-out models of RRP6 and TRF4 genes (Luna *et al.*, 2005, Gavaldá *et al.*, 2013). Accordingly, the instability phenotype in these two mutants is rescued upon overexpression of RNaseH (Wahba *et al.*, 2011), an enzyme that degrades R-loop structures. In previous studies in our group, we detected an increase in R-loop accumulation in H1-TKO cells, coupled to a higher γH2A.X signaling, readout of DNA damage (Almeida *et al.*, 2018). In an analogous way to exosome components mutants, the overexpression of RNaseH decreased the DNA damage response and alleviated the replication defects that are present in these cells, confirming that the transcription-replication conflicts are R-loop dependent and furthering strengthening the possible connection between histone H1 and the exosome complex.

We have detected several transcriptional alterations in H1-TKO cells which could potentially constitute a source of R-loops and, therefore, transcription-replication conflicts. In the first place, transcriptional termination is altered, generating 3' extended RNA that could be linked to a failure in R-loop resection, as happens in mutants of senataxin and *XRN2* (Becherel *et al.*, 2013; Morales *et al.*, 2016). Moreover, there is an accumulation of non-coding transcripts, some of which are known to form R-loop structures *in vivo* (Boque-Sastre *et al.*, 2015).

As discussed before, within the lncRNA group that is accumulated in the chromatin of H1-TKO cells there is a significative fraction of enhancer RNAs. In a recent study, it has been demonstrated that compromising exosome activity produces an accumulation of eRNAs, leading to R-loop formation and DNA damage signaling at enhancer regions (Pefanis *et al.*, 2015). In agreement with this finding, it has been shown that enhancers are fragile sites of the genome, prone to spontaneous mutagenesis (Li and Ovcharenko, 2015), and preferential targets for DNA repair proteins (Sobhy *et al.*, 2019). We propose that reductions in the levels of histone H1 impair the correct exosome-dependent degradation of eRNAs, what would contribute to the observed DNA damage response and replicative defects.

Altogether, these observations help explaining the altered replication initiation landscape we found in H1-TKO cells (Figures 7a, 7c and 7d). Since more than 85% of the mouse genome is transcribed (Hangauer *et al.*, 2013), the altered RNA metabolism and R-loop formation in these cells can occur genome-wide, triggering conflicts with the replication machinery that wouldn't be localized to any specific region of the genome.

The extensive fork stalling caused by these transcription-replication conflicts is likely responsible of the increase in γ-H2A.X signaling (Almeida *et al.*, 2018) and the overactivation of the ATR-Chk1 pathway (Murga *et al.*, 2007) in H1-TKO cells, which would lead to the firing of extra dormant origins to achieve the complete replication of the genome. In agreement with this hypothesis, we have found a hyperphosphorylation of MCM2 in these cells (Figure 6); however, dormant origin firing does not seem sufficient to explain the drastic perturbation of the SNS-seq profile caused by the reduction in the levels of histone H1. In a recent analysis of the replication initiation landscape in two models of extra origin firing (Jodkowska *et al.*, 2019), the resulting SNS-seq profile showed a general increment of origin efficiency, but not the widespread increase of the SNS baseline that we could detect in H1 deficient cells. Thus, other mechanisms would be necessary to account for the accumulation of replication intermediates, like the reinitiation of DNA synthesis at stalled forks by re-priming events. Concerning this possibility, it has been recently found that R-loop replication blocks are bypassed by the cells through the action of the primase-polymerase PrimPol (Svikovic *et al.*, 2019). Primpol is an enzyme which is able to create a short primer after a stalled replication fork and re-start DNA synthesis from it, being then replaced by the replicative DNA polymerase (Martínez-Jiménez *et al.*, 2015). The resulting leading strand would co-purify with the SNS fraction, so these re-priming events would likely contribute to the altered SNS-seq landscape in H1-TKO cells.

CONCLUSIONS

CONCLUSIONES

According to the results presented in this Doctoral Thesis, the following conclusions are drawn:

1. The DNA replication initiation landscape is profoundly disturbed in H1-TKO cells due to the genome-wide presence of replication intermediates.

2. Histone H1 regulates the level of chromatin-bound pre-mRNA synthesized from thousands of coding genes involved in differentiation and development, although this effect has not been detected in the total poly-adenylated fraction.

3. A reduction on the levels of histone H1 triggers a widespread accumulation of unstable non-coding transcripts in chromatin, such as IASs, PROMPTs and lncRNAs. They are nascent RNAs, tethered to chromatin by RNA polymerase II.

4. A significant fraction of the newly detected non-coding transcripts are enhancer RNAs. The accumulation in chromatin of some of them might impair the proper regulation of their partner mRNA during development.

5. Cells with reduced histone H1 levels show defects in several stages of the transcription cycle, including mRNA elongation and termination.

6. Local changes in the nucleosomal landscape or epigenetic marks are not sufficient to explain the altered transcriptional output of H1 deficient cells.

7. The phenotype caused by the reduction of linker histone levels recapitulates the defects triggered by the downregulation of the activity of the nuclear exosome complex, suggesting a possible functional connection between them.

De acuerdo a los resultados presentados en esta Tesis Doctoral, se obtienen las siguientes conclusiones:

1. El paisaje de iniciación de la replicación del DNA está profundamente alterado en células H1-TKO, debido a la presencia de intermediarios replicativos a lo largo de todo el genoma.

2. La histona H1 regula la cantidad de pre-mRNA unido a cromatina de miles de genes codificantes implicados en diferenciación y desarrollo, aunque este efecto no se detecta en la fracción poli-adenilada total.

3. Una reducción de los niveles de histona H1 desencadena una acumulación general de transcritos no codificantes en cromatina, como IASs, PROMPTs y lncRNAs. Estos transcritos son RNAs nacientes, unidos a la cromatina por medio de la RNA polimerasa II.

4. Un porcentaje significativo de los transcritos no codificantes detectados son “enhancer” RNAs. La acumulación en cromatina de algunos de ellos podría dificultar la apropiada regulación de sus RNA mensajeros diana durante el desarrollo.

5. Las células con un contenido reducido de histona H1 muestran defectos en distintas fases del ciclo transcripcional, incluyendo la elongación y la terminación de mRNAs.

6. Las alteraciones transcripcionales en células defectivas para histona H1 no pueden ser explicadas solamente por cambios locales en el paisaje nucleosomal ni en marcas epigenéticas.

7. El fenotipo causado por la disminución de los niveles de H1 recapitula los defectos desencadenados por una reducción de la actividad del complejo exosoma, lo que sugiere una posible conexión funcional entre ellos.

BIBLIOGRAPHY

- Adelman K, Lis JT (2012), 'Promoter-proximal pausing of RNA polymerase II: emerging roles in metazoans', *Nat Rev Genet*, 13(10): 720-731

- Almada AE, Wu X, Kriz AJ, Burge CB, Sharp PA (2013), 'Promoter directionality is controlled by U1 snRNP and polyadenylation signals', *Nature*, 499(7458): 360-363

- Almeida R, Fernández-Justel JM, Santa-María C, Cadoret JC, Cano-Aroca L, Lombraña R, Herranz G, Agresti A, Gómez M (2018), 'Chromatin conformation regulates the coordination between DNA replication and transcription', *Nat Commun*, 23;9(1):1590

- Aloia L, Di Stefano B, Di Croce L (2013), 'Polycomb complexes in stem cells and embryonic development', *Development*, 140(12): 2525-2534

- Andersen PR, Domanski M, Kristiansen MS, Storvall H, Ntini E, Verheggen C, Schein A, Bunkenborg J, Poser I, Hallais M, Sandberg R, Hyman A, LaCava J, Rout MP, Andersen JS, Bertrand E, Jensen TH (2013), 'The human cap-binding complex is functionally connected to the nuclear RNA exosome', *Nat Struct Mol Biol*, 20(12): 1367-1376

- Andersson R, Gebhard C, […], Sandelin A (2014), 'An atlas of active enhancers across human cell types and tissues', *Nature*, 507(7493): 455-461

- Andersson R, Sandelin A, Danko CG (2015), 'A unified architecture of transcriptional regulatory elements', *Trends Genet*, 31(8): 426-433

- Aranda S, Mas G, Di Croce L (2015), 'Regulation of gene transcription by Polycomb proteins', *Sci Adv*, 1(11): e1500737

- Ard R, Allshire RC, Marquardt S (2017), 'Emerging Properties and Functional Consequences of Noncoding Transcription', *Genetics*, 207(2): 357-367

- Arner E, Daub CO, […], Hayashizaki Y (2015), 'Transcribed enhancers lead waves of coordinated transcription in transitioning mammalian cells', *Science*, 347(6225): 1010-1014

- Arya G, Maitra A, Grigoryev SA (2010), 'A structural perspective on the where, how, why, and what of nucleosome positioning', *J Biomol Struct Dyn*, 27(6):803-820

- Azad GK, Ito K, Sailaja BS, Biran A, Nissim-Rafinia M, Yamada Y, Brown DT, Takizawa T, Meshorer E (2018), 'PARP1-dependent eviction of the linker histone H1 mediates immediate early gene expression during neuronal activation', *J Cell Biol*, 217(2): 473-481

- Azvolinsky A, Giresi PG, Lieb JD, Zakian VA (2009), 'Highly transcribed RNA polymerase II genes are impediments to replication fork progression in Saccharomyces cerevisiae', *Mol Cell*, 34(6): 722-734

- Bai L, Morozov AV (2010), 'Gene regulation by nucleosome positioning', *Trends Genet*, 26(11): 476-483

- Barakat TS, Halbritter F, Zhang M, Rendeiro AF, Perenthaler E, Bock C, Chambers I (2018), 'Functional Dissection of the Enhancer Repertoire in Human Embryonic Stem Cells', *Cell Stem Cell*, 23(2): 276-288

- Barboric M, Nissen RM, Kanazawa S, Jabrane-Ferrat N, Peterlin BM (2001), 'NF-κB Binds P-TEFb to Stimulate Transcriptional Elongation by RNA Polymerase II', *Mol Cell*, 8(2): 327-337.

- Barnes CO, Calero M, Malik I, Graham BW, Spahr H, Lin G, Cohen A, Brown IS, Zhang Q, Pullara F, Trakselis MA, Kaplan CD, Calero G (2015), 'Crystal structure of a transcribing RNA Polymerase II complex reveals a complete transcription bubble', *Mol Cell*, 59(2):258-269

- Battaglia S, Lidschreiber M, Baejen C, Torkler P, Vos SM, Cramer P (2017), 'RNA-dependent chromatin association of transcription elongation factors and Pol II CTD kinases', *Elife*, 6. pii: e25637

- Becherel OJ, Yeo AJ, Stellati A, Heng EY, Luff J, Suraweera AM, Woods R, Fleming J, Carrie D, McKinney K, Xu X, Deng C, Lavin MF (2013), 'Senataxin plays an essential role with DNA damage response proteins in meiotic recombination and gene silencing', *PLoS Genet*, 9(4): e1003435

- Bell SP (2002), 'The origin recognition complex: from simple origins to complex functions', *Genes and Dev*, 16(6): 659-672

- Bell SP, Dutta A (2002), 'DNA Replication in Eukaryotic Cells', *Annu Rev Biochem*, 71: 333-374

- Bell SP, Mitchell J, Leber J, Kobayashi R, Stillman B (1995), 'The multidomain structure of Orc1p reveals similarity to regulators of DNA replication and transcriptional silencing', *Cell*, 83(4): 563-568

- Bernstein BE, Mikkelsen TS, Xie X, Kamal M, Huebert DJ, Cuff J, Fry B, Meissner A, Wernig M, Plath K, Jaenisch R, Wagschal A, Feil R, Schreiber SL, Lander ES (2006), 'A bivalent chromatin structure marks key developmental genes in embryonic stem cells', *Cell*, 125(2): 315-326

- Bertone P, Stolc V, Royce TE, Rozowsky JS, Urban AE, Zhu X, Rinn JL, Tongprasit W, Samanta M, Weissman S, Gerstein M, Snyder M (2004), 'Global Identification of Human Transcribed Sequences with Genome Tiling Arrays', *Science*, 306(5705): 2242-2246

- Besnard E, Babled A, Lapasset L, Milhavet O, Parrinello H, Dantec C, Marin JM, Lemaitre JM (2012), 'Unraveling cell type-specific and reprogrammable human replication origin signatures associated with G-quadruplex consensus motifs', *Nat Struct Mol Biol*, 19(8): 837-844

- Birney E, Stamatoyannopoulos JA, Dutta A, […], de Jong PJ (2007), 'Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project', *Nature*, 447(7146): 799-816

- Blinka S, Reimer Jr. MH, Pulakanti K, Rao S (2016), 'Super-Enhancers at the Nanog Locus Differentially Regulate Neighboring Pluripotency-Associated Genes', *Cell Rep*, 17(1):19-28

- Blow JJ, Ge XQ, Jackson DA (2011), 'How dormant origins promote complete genome replication', *Trends Biochem Sci*, 36(8): 405-414

- Boque-Sastre R, Soler M, Oliveira-Mateos C, Portela A, Moutinho C, Sayols S, Villanueva A, Esteller M, Guil S (2015), 'Head-to-head antisense transcription and R-loop formation promotes transcriptional activation', *Proc Natl Acad Sci USA*, 112(18): 5785-5790

- Bousquet-Antonelli C, Presutti C, Tollervey D (2000), 'Identification of a regulated pathway for nuclear pre-mRNA turnover', *Cell*, 102(6): 765-75

- Branzei D (2011), 'Ubiquitin family modifications and template switching', *FEBS Lett*, 585(18): 2810-2817

- Braunschweig U, Hogan GJ, Pagie L, van Steensel B (2009), 'Histone H1 binding is inhibited by histone variant H3.3', *EMBO J*, 28(23): 3635-3645

- Bresnick EH, Bustin M, Marsaud V, Richard-Foy H, Hager GL (1992), 'The transcriptionally-active MMTV promoter is depleted of histone H1', *Nucleic Acids Res*, 20(2):273-278

- Bühler M (2009), 'RNA turnover and chromatin-dependent gene silencing', *Chromosoma*, 118(2): 141-151

- Bulger M, Groudine M (2011), 'Functional and Mechanistic Diversity of Distal Transcription Enhancers', *Cell*, 144(3): 327-339

- Cadoret JC, Meisch F, Hassan-Zadeh V, Luyten I, Guillet C, Duret L, Quesneville H, Prioleau MN (2008), 'Genome-wide studies highlight indirect links between human replication origins and gene regulation', *Proc Natl Acad Sci USA*, 105(41): 15837–15842

- Cao K, Lailler N, Zhang Y, Kumar A, Uppal K, Liu Z, Lee EK, Wu H, Medrzycki M, Pan C, Ho PY, Cooper GP Jr, Dong X, Bock C, Bouhassira EE, Fan Y (2013), 'High-resolution mapping of h1 linker histone variants in embryonic stem cells', *PLoS Genet*, 9(4): e1003417

- Catarino RR, Stark A (2018), 'Assessing sufficiency and necessity of enhancer activities for gene expression and the mechanisms of transcription activation', *Genes Dev*, 32(3-4): 202-223

- Cayrou C, Coulombe P, Vigneron A, Stanojcic S, Ganier O, Peiffer I, Rivals E, Puy A, Laurent-Chabalier S, Desprat R, Mechali M (2011), 'Genome-scale analysis of metazoan replication origins reveals their organization in specific but flexible sites defined by conserved features', *Genome Res*, 21(9): 1438–1449

- Cayrou C, Ballester B, Peiffer I, Fenouil R, Coulombe P, Andrau JC, van Helden J, Mechali M (2015), 'The chromatin environment shapes DNA replication origin organization and defines origin classes', *Genome Res*, 25(12): 1873–1885

- Chakraborty P, Huang JTJ, Hiom K (2018), 'DHX9 helicase promotes R-loop formation in cells with impaired RNA splicing', *Nat Commun*, 9(1): 4346

- Changolkar LN, Pehrson JR (2006), 'macroH2A1 histone variants are depleted on active genes but concentrated on the inactive X chromosome', *Mol Cell Biol*, 26(12): 4410-4420

- Chen C, Lim HH, Shi J, Tamura S, Maeshima K, Surana U, Gan L (2016), 'Budding yeast chromatin is dispersed in a crowded nucleoplasm in vivo', *Mol Biol Cell*, 27(21): 3357-3368

- Chen FX, Woodfin AR, Gardini A, Rickels RA, Marshall SA, Smith ER, Shiekhattar R, Shilatifard A (2015), 'PAF1, a Molecular Regulator of Promoter-Proximal Pausing by RNA Polymerase II', *Cell*, 162(5):1003-1015

- Chen FX, Xie P, Collings CK, Cao K, Aoi Y, Marshall SA, Rendleman EJ, Ugarenko M, Ozark PA, Zhang A, Shiekhattar R, Smith ER, Zhang MQ, Shilatifard A (2017), 'PAF1 regulation of promoter-proximal pause release via enhancer activation', *Science*, 357(6357): 1294-1298

- Chen FX, Smith ER, Shilatifard A (2018), 'Born to run: control of transcription elongation by RNA polymerase II', *Nat Rev Mol Cell Biol*, 19(7): 464-478

- Chen K, Xi Y, Pan X, Li Z, Kaestner K, Tyler J, Dent S, He X, Li W (2013), 'DANPOS: dynamic analysis of nucleosome position and occupancy by sequencing', *Genome Res*, 23(2): 341-351

- Chen S, Bell SP (2011), 'CDK prevents Mcm2–7 helicase loading by inhibiting Cdt1 interaction with Orc6', *Genes Dev*, 25(4): 363–372

- Chen YH, Keegan S, Kahli M, Tonzi P, Fenyö D, Huang TT, Smith DJ (2019), 'Transcription shapes DNA replication initiation and termination in human cells', *Nat Struct Mol Biol*, 26(1): 67-77

- Cheng GH, Nandi A, Clerk S, Skoultchi AI (1989), 'Different 3'-end processing produces two independently regulated mRNAs from a single H1 histone gene', *Proc Natl Acad Sci USA*, 86(18): 7002-7006

- Cheng J, Blum R, Bowman C, Hu D, Shilatifard A, Shen S, Dynlacht BD (2014), 'A role for H3K4 monomethylation in gene repression and partitioning of chromatin readers', *Mol Cell*, 53(6): 979-992

- Chiu AC, Suzuki HI, Wu X, Mahat DB, Kriz AJ, Sharp PA (2018), 'Transcriptional Pause Sites Delineate Stable Nucleosome-Associated Premature Polyadenylation Suppressed by U1 snRNP', *Mol Cell*, 69(4): 648-663

- Church MC, Fleming AB (2018), 'A role for histone acetylation in regulating transcription elongation', *Transcription*, 9(4): 225-232

- Clapier CR, Iwasa J, Cairns BR, Peterson CL (2017), 'Mechanisms of action and regulation of ATP-dependent chromatin-remodelling complexes', *Nat Rev Mol Cell Biol*, 18(7): 407-422

- Cobb JA, Bjergbaek L, Shimada K, Frei C, Gasser SM (2003), 'DNA polymerase stabilization at stalled replication forks requires Mec1 and the RecQ helicase Sgs1', *EMBO J*, 22(16): 4325–4336

- Colin J, Candelli T, Porrua O, Boulay J, Zhu C, Lacroute F, Steinmetz LM, Libri D (2014), 'Roadblock termination by reb1p restricts cryptic and readthrough transcription', *Mol Cell*, 56(5): 667-680

- Courtot L, Hoffmann JS, Bergoglio V (2018), 'The Protective Role of Dormant Origins in Response to Replicative Stress', *Int J Mol Sci*, 19(11): 3569

- Creyghton MP, Cheng AW, Welstead GG, Kooistra T, Carey BW, Steine EJ, Hanna J, Lodato MA, Frampton GM, Sharp PA, Boyer LA, Young RA, Jaenisch R (2010), 'Histone H3K27ac separates active from poised enhancers and predicts developmental state', *Proc Natl Acad Sci USA*, 107(50): 21931-21936

- Crossley MP, Bocek M, Cimprich KA (2019), 'R-Loops as Cellular Regulators and Genomic Threats', *Mol Cell*, 73(3): 398-411

- Datta A, Jinks-Robertson S (1995), 'Association of increased spontaneous mutation rates with high levels of transcription in yeast', *Science*, 268(5217): 1616-1619

- Dave K, Sur I, Yan J, Zhang J, Kaasinen E, Zhong F, Blaas L, Li X, Kharazi S, Gustafsson C, De Paepe A, Mansson R, Taipale J (2017), 'Mice deficient of Myc super-enhancer region reveal differential control mechanism between normal and pathological growth', *Elife*, 6. pii: e23382

- De Santa F, Barozzi I, Mietton F, Ghisletti S, Polletti S, Tusi BK, Muller H, Ragoussis J, Wei CL, Natoli G (2010), 'A Large Fraction of Extragenic RNA Pol II Transcription Sites Overlap Enhancers', *PLoS Biol*, 8(5): e1000384

- Delgado S, Gómez M, Bird A, Antequera F (1998), 'Initiation of DNA replication at CpG islands in mammalian chromosomes', *EMBO J*, 17(8): 2426–2435

- Dellino GI, Cittaro D, Piccioni R, Luzi L, Banfi S, Segalla S, Cesaroni M, Mendoza-Maldonado R, Giacca M, Pelicci PG (2013), 'Genome-wide mapping of human DNA-replication origins: Levels of transcription at ORC1 sites regulate origin selection and replication timing', *Genome Res*, 23(1): 1–11

- DePamphilis ML (1993), 'Origins of DNA replication in metazoan chromosomes', *J Biol Chem*, 268(1):1-4

- DesJarlais R, Tummino PJ (2016), 'Role of Histone-Modifying Enzymes and Their Complexes in Regulation of Chromatin Biology', *Biochemistry*, 55(11): 1584-1599

- Diffley JFX (2004), 'Regulation of Early Events in Chromosome Replication', *Curr Biol*, 14(18): R778-786.

- Djebali S, Davis CA, […], Gingeras TR (2012), 'Landscape of transcription in human cells', *Nature*, 489(7414): 101-108

- Dong X, Weng Z (2013), 'The correlation between histone modifications and gene expression', *Epigenomics*, 5(2): 113-116.

- Dujardin G, Lafaille C, de la Mata M, Le Jossic-Corcos C, Corcos L, Kornblihtt AR (2014), 'How Slow RNA Polymerase II Elongation Favors Alternative Exon Skipping', *Mol Cell*, 54(4): 683-690

- Duquette ML, Handa P, Vincent JA, Taylor AF, Maizels N (2004), 'Intracellular transcription of G-rich DNAs induces formation of G-loops, novel structures containing G4 DNA', *Genes Dev*, 18(13): 1618–1629

- Eaton ML, Galani K, Kang S, Bell SP, MacAlpine DM (2010), 'Conserved nucleosome positioning defines replication origins', Genes and Dev, 24(8): 748–753

- Ellison MA, Lederer AR, Warner MH, Mavrich TN, Raupach EA, Heisler LE, Nislow C, Lee MT, Arndt KM (2019), 'The Paf1 Complex Broadly Impacts the Transcriptome of Saccharomyces cerevisiae', *Genetics*, 212(3): 711-728

- Elvers I, Johansson F, Groth P, Erixon K, Helleday T (2011), 'UV stalled replication forks restart by re-priming in human fibroblasts', *Nucleic Acid Res*, 39(16): 7049–7057

- Fan Y, Nikitina T, Morin-Kensicki EM, Zhao J, Magnuson TR, Woodcock CL, Skoultchi AI (2003), 'H1 linker histones are essential for mouse development and affect nucleosome spacing in vivo', *Mol Cell Biol*, 23(13): 4559-4572

- Fan Y, Nikitina T, Zhao J, Fleury TJ, Bhattacharyya R, Bouhassira EE, Stein A, Woodcock CL, Skoultchi AI (2005), 'Histone H1 depletion in mammals alters global chromatin structure but causes specific changes in gene regulation', *Cell*, 123(7): 1199-1212

- Fan Y, Sirotkin A, Russell RG, Ayala J, Skoultchi AI (2001), 'Individual somatic H1 subtypes are dispensable for mouse development even in mice lacking the H1(0) replacement subtype', *Mol Cell Biol*, 21(23): 7933-7943

- Felipe-Abrio I, Lafuente-Barquero J, García-Rubio ML, Aguilera A (2015), 'RNA polymerase II contributes to preventing transcription-mediated replication fork stalls', *EMBO J*, 34(2): 236–250

- Felsenfeld G, Groudine M (2003), 'Controlling the double helix', *Nature*, 421(6921): 448-453

- Feng Y, Vlassis A, Roques C, Lalonde ME, González-Aguilera C, Lambert JP, Lee SB, Zhao X, Alabert C, Johansen JV, Paquet E, Yang XJ, Gingras AC, Côté J, Groth A (2015), 'BRPF3-HBO1 regulates replication origin activation and histone H3K14 acetylation', *EMBO J*, 35: 176-192

- Fischl H, Howe FS, Furger A, Mellor J (2017), 'Paf1 Has Distinct Roles in Transcription Elongation and Differential Transcript Fate', *Mol Cell*, 65(4): 685-698

- Flanagan TW, Brown DT (2016), 'Molecular dynamics of histone H1', *Biochim Biophys Acta*, 1859(3): 468-475

- Flynn RA, Almada AE, Zamudio JR, Sharp PA (2011), 'Antisense RNA polymerase II divergent transcripts are P-TEFb dependent and substrates for the RNA exosome', *Proc Natl Acad Sci USA*, 108(26): 10460-10465

- Fong N, Kim H, Zhou Y, Ji X, Qiu J, Saldi T, Diener K, Jones K, Fu XD, Bentley DL (2014), 'Pre-mRNA splicing is facilitated by an optimal RNA polymerase II elongation rate', *Genes Dev*, 28(23): 2663-2676.

- Francis LI, Randell JCW, Takara TJ, Uchima L, Bell SP (2009), 'Incorporation into the prereplicative complex activates the Mcm2–7 helicase for Cdc7–Dbf4 phosphorylation', *Genes Dev*, 23(5):643-654

- French S (1992), 'Consequences of replication fork movement through transcription units in vivo', *Science*, 258(5086): 1362-1365

- Friedman N, Rando OJ (2015), 'Epigenomics and the structure of the living genome', *Genome Res*, 25(10):1482-1490

- Fussner E, Strauss M, Djuric U, Li R, Ahmed K, Hart M, Ellis J, Bazett-Jones DP (2012), 'Open and closed domains in the mouse genome are configured as 10-nm chromatin fibres', *EMBO Rep*, 13(11): 992-996

- Gaertner B, Zeitlinger J (2014), 'RNA polymerase II pausing during development', *Development*, 141(6): 1179-1183

- García-Benítez F, Gaillard H, Aguilera A (2017), 'Physical proximity of chromatin to nuclear pores prevents harmful R loop accumulation contributing to maintain genome stability', *Proc Natl Acad Sci USA*, 114(41): 10942-10947

- Gavaldá S, Gallardo M, Luna R, Aguilera A (2013), 'R-loop mediated transcription-associated recombination in trf4Δ mutants reveals new links between RNA surveillance and genome integrity', *PLoS One*, 8(6): e65541

- Ge XQ, Blow JJ (2010), 'Chk1 inhibits replication factory activation but allows dormant origin firing in existing factories', *J Cell Biol*, 191(7):1285-1297

- Ge XQ, Jackson DA, Blow JJ (2007), 'Dormant origins licensed by excess Mcm2–7 are required for human cells to survive replicative stress', *Genes Dev*, 21(24): 3331–3341

- Geeven G, Zhu Y, Kim BJ, Bartholdy BA, Yang SM, Macfarlan TS, Gifford WD, Pfaff SL, Verstegen MJ, Pinto H, Vermunt MW, Creyghton MP, Wijchers PJ, Stamatoyannopoulos JA, Skoultchi AI, de Laat W (2015), 'Local compartment changes and regulatory landscape alterations in histone H1-depleted cells', *Genome Biol*, 16: 289

- Ghavi-Helm Y, Klein FA, Pakozdi T, Ciglar L, Noordermeer D, Huber W, Furlong EEM (2014), 'Enhancer loops appear stable during development and are associated with paused polymerase', *Nature*, 512(7512): 96-100

- Giaimo BD, Ferrante F, Herchenröther A, Hake SB, Borggrefe T (2019), 'The histone variant H2A.Z in gene regulation', *Epigenetics Chromatin*, 12(1):37

- Gilchrist DA, Nechaev S, Lee C, Kumar S, Ghosh B, Collins JB, Li L, Gilmour DS, Adelman K (2008), 'NELF-mediated stalling of Pol II can enhance gene expression by blocking promoter-proximal nucleosome assembly', *Genes Dev*, 22(14): 1921-1933

- González S, García A, Vázquez E, Serrano R, Sánchez M, Quintales L, Antequera F (2016), 'Nucleosomal signatures impose nucleosome positioning in coding and noncoding sequences in the genome', *Genome Res*, 26(11): 1532-1543

- Gräff J, Tsai LH (2013), 'Histone acetylation: molecular mnemonics on the chromatin', *Nat Rev Neurosci*, 14(2): 97-111

- Grosso AR, Leite AP, Carvalho S, Matos MR, Martins FB, Vítor AC, Desterro JM, Carmo-Fonseca M, de Almeida SF (2015), 'Pervasive transcription read-through promotes aberrant expression of oncogenes and RNA chimeras in renal carcinoma', *Elife*, 4. pii: e09214

- Gu B, Lee MG (2013), 'Histone H3 lysine 4 methyltransferases and demethylases in self-renewal and differentiation of stem cells', *Cell Biosci*, 3(1):39

- Gudipati RK, Neil H, Feuerbach F, Malabat C, Jacquier A (2012), 'The yeast RPL9B gene is regulated by modulation between two modes of transcription termination', *EMBO J*, 31(10): 2427-2437

- Hah N, Benner C, Chong LW, Yu RT, Downes M, Evans RM (2015), 'Inflammation-sensitive super enhancers form domains of coordinately regulated enhancer RNAs', *Proc Natl Acad Sci USA*, 112(3): E297-302

- Hah N, Murakami S, Nagari A, Danko CG, Kraus WL (2013), 'Enhancer transcripts mark active estrogen receptor binding sites', *Genome Res*, 23(8):1210-1223

- Hangauer MJ, Vaughn IW, McManus MT (2013), 'Pervasive Transcription of the Human Genome Produces Thousands of Previously Unidentified Long Intergenic Noncoding RNAs', *PLoS Genet*, 9(6): e1003569

- Harshman SW, Young NL, Parthun MR, Freitas MA (2013), 'H1 histones: current perspectives and challenges', *Nucleic Acids Res*, 41(21): 9593-9609

- Hashimoto H, Takami Y, Sonoda E, Iwasaki T, Iwano H, Tachibana M, Takeda S, Nakayama T, Kimura H, Shinkai Y (2010), 'Histone H1 null vertebrate cells exhibit altered nucleosome architecture', *Nucleic Acids Res*, 38(11): 3533-3545

- Hassan-Zadeh V, Chilaka V, Cadoret JC, Ma MKW, Boggetto N, West A, Prioleau MN (2012), 'USF Binding Sequences from the HS4 Insulator Element Impose Early Replication Timing on a Vertebrate Replicator', *PLoS Biol*, 10(3): e1001277

- Heinz S, Benner C, Spann N, Bertolino E, Lin YC, Laslo P, Cheng JX, Murre C, Singh H, Glass CK (2010), 'Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities', *Mol Cell*, 38(4):576-589

- Heller RC, Kang S, Lam WM, Chen S, Chan CS, Bell SP (2011), 'Eukaryotic Origin-Dependent DNA Replication In Vitro Reveals Sequential Action of DDK and S-CDK Kinases', *Cell*, 146(1): 80-91

- Hergeth SP, Schneider R (2015), 'The H1 linker histones: multifunctional proteins beyond the nucleosomal core particle', *EMBO Rep*, 16(11): 1439-1453

- Hiratani I, Ryba T, Itoh M, Yokochi T, Schwaiger M, Chang CW, Lyou Y, Townes TM, Schübeler D, Gilbert DM (2008), 'Global reorganization of replication domains during embryonic stem cell differentiation', *PLoS Biol*, 6(10): e245

- Ho Y, Elefant F, Liebhaber SA, Cooke NE (2006), 'Locus Control Region Transcription Plays an Active Role in Long-Range Gene Activation', *Mol Cell*, 23(3):365-375

- Hocine S, Singer RH, Grünwald D (2010), 'RNA Processing and Export', *Cold Spring Harb Perspect Biol*, 2(12): a000752

- Hodroj D, Serhal K, Maiorano D (2017), 'Ddx19 links mRNA nuclear export with progression of transcription and replication and suppresses genomic instability upon DNA damage in proliferating cells', *Nucleus*, 8(5): 489-495

- Hoggard T, Shor E, Müller CA, Nieduszynski CA, Fox CA (2013), 'A Link between ORC-Origin Binding Mechanisms and Origin Activation Time Revealed in Budding Yeast', *PLoS Genet*, 9(9): e1003798

- Hou L, Wang Y, Liu Y, Zhang N, Shamovsky I, Nudler E, Tian B, Dynlacht BD (2019), 'Paf1C regulates RNA polymerase II progression by modulating elongation rate', *Proc Natl Acad Sci USA*, 116(29): 14583-14592

- Howman EV, Fowler KJ, Newson AJ, Redward S, MacDonald AC, Kalitsis P, Choo KH (2000), 'Early disruption of centromeric chromatin organization in centromere protein A (Cenpa) null mice', *Proc Natl Acad Sci USA*, 97(3):1148-1153

- Hrossova D, Sikorsky T, Potesil D, Bartosovic M, Pasulka J, Zdrahal Z, Stefl R, Vanacova S (2015), 'RBM7 subunit of the NEXT complex binds U-rich sequences and targets 3'-end extended forms of snRNAs', *Nucleic Acids Res*, 43(8): 4236-4248

- Hsieh FK, Kulaeva OI, Patel SS, Dyer PN, Luger K, Reinberg D, Studitsky VM (2013), 'Histone chaperone FACT action during transcription through chromatin by RNA polymerase II', *Proc Natl Acad Sci USA*, 110(19): 7654-7659

- Hsin JP, Manley JL (2012), 'The RNA polymerase II CTD coordinates transcription and RNA processing', *Genes Dev*, 26(19): 2119-2137

- Hu J, Gu L, Ye Y, Zheng M, Xu Z, Lin J, Du Y, Tian M, Luo L, Wang B, Zhang X, Weng Z, Jiang C (2018), 'Dynamic placement of the linker histone H1 associated with nucleosome arrangement and gene transcription in early Drosophila embryonic development', *Cell Death Dis*, 9(7):765

- Ibarra A, Schwob E, Méndez J (2008), 'Excess MCM proteins protect human cells from replicative stress by licensing backup origins of replication', *Proc Natl Acad Sci USA*, 105(26):8956-8961

- Ilott NE, Heward JA, Roux B, Tsitsiou E, Fenwick PS, Lenzi L, Goodhead I, Hertz-Fowler C, Heger A, Hall N, Donnelly LE, Sims D, Lindsay MA (2014), 'Long non-coding RNAs and enhancer RNAs regulate the lipopolysaccharide-induced inflammatory response in human monocytes', *Nat Commun*, 5: 3979

- Ioshikhes I, Bolshoy A, Derenshteyn K, Borodovsky M, Trifonov EN (1996), 'Nucleosome DNA sequence pattern revealed by multiple alignment of experimentally mapped sequences', *J Mol Biol*, 262(2): 129-139

- Izquierdo-Bouldstridge A, Bustillos A, Bonet-Costa C, Aribau-Miralbés P, García-Gomis D, Dabad M, Esteve-Codina A, Pascual-Reguant L, Peiró S, Esteller M, Murtha M, Millán-Ariño L, Jordan A (2017), 'Histone H1 depletion triggers an interferon response in cancer cells via activation of heterochromatic repeats', *Nucleic Acids Res*, 45(20): 11622-11642

- Izzo A, Kamieniarz-Gdula K, Ramírez F, Noureen N, Kind J, Manke T, van Steensel B, Schneider R (2013), 'The genomic landscape of the somatic linker histone subtypes H1.1 to H1.5 in human cells', *Cell Rep*, 3(6): 2142-2154

- Jaquier A (2009), 'The complex eukaryotic transcriptome: unexpected pervasive transcription and novel small RNAs', *Nat Rev Genet*, 10(12): 833-844

- Jiang C, Pugh F (2009), 'Nucleosome positioning and gene regulation: advances through genomics', *Nat Rev Genet*, 10(3): 161-72

- Jimeno-González S, Ceballos-Chávez M, Reyes JC (2015a), 'A positioned +1 nucleosome enhances promoter-proximal pausing', *Nucleic Acids Res*, 43(6): 3068-3078

- Jimeno-González S, Payán-Bravo L, Muñoz-Cabello AM, Guijo M, Gutiérrez G, Prado F, Reyes JC (2015b), 'Defective histone supply causes changes in RNA polymerase II elongation rate and cotranscriptional pre-mRNA splicing', *Proc Natl Acad Sci USA*, 112(48): 14840-14845

- Jin C, Zang C, Wei G, Cui K, Peng W, Zhao K, Felsenfeld G (2009), 'H3.3/H2A.Z double variant-containing nucleosomes mark 'nucleosome-free regions' of active promoters and other regulatory regions', *Nat Genet*, 41(8): 941-945

- Jodkowska K, Pancaldi V, Almeida R, Rigau M, Graña-Castro O, Fernández-Justel JM, Rodríguez-Acebes S, Rubio-Camarillo M, Carrillo de Santa Pau E, Pisano D, Al-Shahrour F, Valencia A, Gómez M, Méndez J (2019), 'Three-dimensional connectivity

and chromatin environment mediate the activation efficiency of mammalian DNA replication origins', *bioRxiv,* doi:10.1101/644971

- Jonkers I, Lis JT (2015), 'Getting up to speed with transcription elongation by RNA polymerase II', *Nat Rev Mol Cell Biol*, 16(3): 167-177

- Juan D, Perner J, Carrillo de Santa Pau E, Marsili S, Ochoa D, Chung HR, Vingron M, Rico D, Valencia A (2016), 'Epigenomic Co-localization and Co-evolution Reveal a Key Role for 5hmC as a Communication Hub in the Chromatin Network of ESCs', *Cell Rep*, 14(5): 1246-1257

- Kalashnikova AA, Winkler DD, McBryant SJ, Henderson RK, Herman JA, DeLuca JG, Luger K, Prenni JE, Hansen JC (2013), 'Linker histone H1.0 interacts with an extensive network of proteins found in the nucleolus', *Nucleic Acids Res*, 41(7): 4026-4035

- Kamieniarz K, Izzo A, Dundr M, Tropberger P, Ozretic L, Kirfel J, Scheer E, Tropel P, Wisniewski JR, Tora L, Viville S, Buettner R, Schneider R (2012), 'A dual role of linker histone H1.4 Lys 34 acetylation in transcriptional activation', *Genes Dev*, 26(8): 797-802

- Kamieniearz-Gdula K, Gdula MR, Panser K, Nojima T, Monks J, Wiśniewski JR, Riepsaame J, Brockdorff N, Pauli A, Proudfoot NJ (2019), 'Selective Roles of Vertebrate PCF11 in Premature and Full-Length Transcript Termination', *Mol Cell*, 74(1):158-172

- Kawabata T, Luebben SW, Yamaguchi S, Ilves I, Matise I, Buske T, Botchan MR, Shima N (2011), 'Stalled fork rescue via dormant replication origins in unchallenged S phase promotes proper chromosome segregation and tumor suppression', *Mol Cell*, 41(5): 543-553

- Khan A, Zhang X (2016), 'dbSUPER: a database of super-enhancers in mouse and human genome', *Nucleic Acids Res*, 44(D1):D164-71

- Kim D, Pertea G, Trapnell C, Pimentel H, Kelley R, Salzberg SL (2013), 'TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions', *Genome Biol*, 14(4):R36

- Kim JM, Kim K, Punj V, Liang G, Ulmer TS, Lu W, An W (2015), 'Linker histone H1.2 establishes chromatin compaction and gene silencing through recognition of H3K27me3', *Sci Rep*, 5: 16714

- Kim K, Choi J, Heo K, Kim H, Levens D, Kohno K, Johnson EM, Brock HW, An W (2008), 'Isolation and characterization of a novel H1.2 complex that acts as a repressor of p53-mediated transcription', *J Biol Chem*, 283(14): 9113-9126

- Kim K, Lee B, Kim J, Choi J, Kim JM, Xiong Y, Roeder RG, An W (2013), 'Linker Histone H1.2 cooperates with Cul4A and PAF1 to drive H4K31 ubiquitylation-mediated transactivation', *Cell Rep*, 5(6): 1690-1703

- Kim TK, Hemberg M, Gray JM, Costa AM, Bear DM, Wu J, Harmin DA, Laptewicz M, Barbara-Haley K, Kuersten S, Markenscoff-Papadimitriou E, Kuhl D, Bito H, Worley PF, Kreiman G, Greenberg ME (2010), 'Widespread transcription at neuronal activity-regulated enhancers', *Nature*, 465(7295): 182-187

- King IF, Yandava CN, Mabb AM, Hsiao JS, Huang HS, Pearson BL, Calabrese JM, Starmer J, Parker JS, Magnuson T, Chamberlain SJ, Philpot BD, Zylka MJ (2013), 'Topoisomerases facilitate transcription of long genes linked to autism', *Nature*, 501(7465): 58-62

- Knott SR, Peace JM, Ostrow AZ, Viggiani CJ, Tavaré S, Aparicio OM (2012), 'Forkhead Transcription Factors Establish Origin Timing and Long-Range Clustering in S. cerevisiae', *Cell*, 148(1-2): 99-111

- Knott SR, Viggiani CJ, Tavaré S, Aparicio OM (2009), 'Genome-wide replication profiles indicate an expansive role for Rpd3L in regulating replication initiation timing or efficiency, and reveal genomic loci of Rpd3 function in Saccharomyces cerevisiae', *Genes Dev*, 23(9): 1077-1090

- Kornberg RD (1974), 'Chromatin structure: a repeating unit of histones and DNA', *Science*, 184(4139): 868-871

- Kotake Y, Goto T, Naemura M, Inoue Y, Okamoto H, Tahara K (2017), 'Long Noncoding RNA PANDA Positively Regulates Proliferation of Osteosarcoma Cells', *Anticancer Res*, 37(1): 81-85.

- Kowalski A (2016), 'Nuclear and nucleolar activity of linker histone variant H1.0', *Cell Mol Biol Lett*, 21:15

- Krishnakumar R, Gamble MJ, Frizzell KM, Berrocal JG, Kininis M, Kraus WL (2008), 'Reciprocal binding of PARP-1 and histone H1 at promoters specifies transcriptional outcomes', *Science*, 319(5864): 819-821

- Kuleshov MV, Jones MR, Rouillard AD, Fernandez NF, Duan Q, Wang Z, Koplev S, Jenkins SL, Jagodnik KM, Lachmann A, McDermott MG, Monteiro CD, Gundersen GW, Ma'ayan A (2016), 'Enrichr: a comprehensive gene set enrichment analysis web server 2016 update', *Nucleic Acids Res*, 44(W1): W90-97

- Kuo AJ, Song J, Cheung P, Ishibe-Murakami S, Yamazoe S, Chen JK, Patel DJ, Gozani O (2012), 'The BAH domain of ORC1 links H4K20me2 to DNA replication licensing and Meier–Gorlin syndrome', *Nature*, 84(7392): 115-119

- Lai F, Orom UA, Cesaroni M, Beringer M, Taatjes DJ, Blobel GA, Shiekhattar R (2013), 'Activating RNAs associate with Mediator to enhance chromatin architecture and transcription', *Nature*, 494(7438): 497-501

- Latgé G, Poulet C, Bours V, Josse C, Jerusalem G (2018), 'Natural Antisense Transcripts: Molecular Mechanisms and Implications in Breast Cancers', *Int J Mol Sci*, 2;19(1)

- Lauberth SM, Nakayama T, Wu X, Ferris AL, Tang Z, Hughes SH, Roeder RG (2014), 'H3K4me3 interactions with TAF3 regulate preinitiation complex assembly and selective gene activation', *Cell*, 152(5): 1021-1036

- Lee S, Seo HH, Lee CY, Lee J, Shin S, Kim SW, Lim S, Hwang KC (2017), 'Human Long Noncoding RNA Regulation of Stem Cell Potency and Differentiation', *Stem Cells Int*, 2017: 6374504

- Lemay JF, Larochelle M, Marguerat S, Atkinson S, Bähler J, Bachand F (2014), 'The RNA exosome promotes transcription termination of backtracked RNA polymerase II', *Nat Struct Mol Biol*, 21(10): 919-26

- Lemieux C, Marguerat S, Lafontaine J, Barbezier N, Bähler J, Bachand F (2011), 'A Pre-mRNA degradation pathway that selectively targets intron-containing genes requires the nuclear poly(A)-binding protein', *Mol Cell*, 44(1): 108-119

- Lever MA, Th'ng JP, Sun X, Hendzel MJ (2000), 'Rapid exchange of histone H1.1 on chromatin in living human cells', *Nature*, 408(6814): 873-876

- Levine M (2010), 'Transcriptional Enhancers in Animal Development and Evolution', *Curr Biol*, 20(17): R754-763

- Li S, Ovcharenko I (2015), 'Human Enhancers Are Fragile and Prone to Deactivating Mutations', *Mol Biol Evol*, 32(8): 2161-2180

- Li W, Notani D, Ma Q, Tanasa B, Nunez E, Chen AY, Merkurjev D, Zhang J, Ohgi K, Song X, Oh S, Kim HS, Glass CK, Rosenfeld MG (2013), 'Functional roles of enhancer RNAs for oestrogen-dependent transcriptional activation', *Nature*, 498(7455): 516-520

- Li X, Manley JL (2005), 'Inactivation of the SR protein splicing factor ASF/SF2 results in genomic instability', *Cell*, 122(3): 365-378

- Li Y, Rivera CM, Ishii H, Jin F, Selvaraj S, Lee AY, Dixon JR, Ren B (2014), 'CRISPR reveals a distal super-enhancer required for Sox2 expression in mouse embryonic stem cells', *PLoS One*, 9(12): e114485

- Liang K, Gao X, Gilmore JM, Florens L, Washburn MP, Smith E, Shilatifard A (2015), 'Characterization of human cyclin-dependent kinase 12 (CDK12) and CDK13 complexes in C-terminal domain phosphorylation, gene transcription, and RNA processing', *Mol Cell Biol*, 35(6): 928-938

- Liao J, He Y, Szabó PE (2013), 'The Pou5f1 distal enhancer is sufficient to drive Pou5f1 promoter-EGFP expression in embryonic stem cells', *Int J Dev Biol*, 57(9-10): 725-729

- Lin D, Hiron TK, O'Callaghan CA (2018), 'Intragenic transcriptional interference regulates the human immune ligand MICA', *EMBO J*, 15;37(10)

- Lin Q, Sirotkin A, Skoultchi AI (2000), 'Normal spermatogenesis in mice lacking the testis-specific linker histone H1t', *Mol Cell Biol*, 20(6): 2122-2128

- Ling J, Ainol L, Zhang L, Yu X, Pi W, Tuan D (2004), 'HS2 Enhancer Function Is Blocked by a Transcriptional Terminator Inserted between the Enhancer and the Promoter', *J Biol Chem*, 279(49):51704-51713

- Lipford JR, Bell SP (2001), 'Nucleosomes positioned by ORC facilitate the initiation of DNA replication', *Mol Cell*, 7(1):21-30

- Liu LF, Wang JC (1987), 'Supercoiling of the DNA template during transcription', *Proc Natl Acad Sci USA*, 84(20): 7024–7027

- Lloret-Llinares M, Karadoulama E, Chen Y, Wojenski LA, Villafano GJ, Bornholdt J, Andersson R, Core L, Sandelin A, Jensen TH (2018), 'The RNA exosome contributes to gene expression regulation during stem cell differentiation', *Nucleic Acids Res*, 46(21): 11502-11513

- Lloret-Llinares M, Mapendano CK, Martlev LH, Lykke-Andersen S, Jensen TH (2016), 'Relationships between PROMPT and gene expression', *RNA Biol*, 13(1):6-14

- Lombraña R, Almeida R, Revuelta I, Madeira S, Herranz G, Saiz N, Bastolla U, Gómez M (2013), 'High-resolution analysis of DNA synthesis start sites and nucleosome architecture at efficient mammalian replication origins', *EMBO J*, 32(19): 2631–2644

- Loomis EW, Sanz LA, Chedin F, Hagerman PJ (2014), 'Transcription-Associated R-Loop Formation across the Human FMR1 CGG-Repeat Region', *PLoS Genet*, 10(4):e1004294

- Lopes M, Foiani M, Sogo JM (2006), 'Multiple Mechanisms Control Chromosome Integrity after Replication Fork Uncoupling and Restart at Irreparable UV Lesions', *Mol Cell*, 21(1):15-27

- Lopes M, Cotta-Ramusino, Pellicioli A, Liberi G, Plevani P, Muzi-Falconi M, Newlon CS, Foiani M (2001), 'The DNA replication checkpoint response stabilizes stalled replication forks', *Nature*, 412(6846): 557-561

- Love MI, Huber W, Anders S (2014), 'Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2', *Genome Biol*, 15(12): 550.

- Lu X, Wontakal SN, Emelyanov AV, Morcillo P, Konev AY, Fyodorov DV, Skoultchi AI (2009), 'Linker histone H1 is essential for Drosophila development, the establishment of pericentric heterochromatin, and a normal polytene chromosome structure', *Genes Dev*, 23(4): 452-465

- Lubas M, Andersen PR, Schein A, Dziembowski A, Kudla G, Jensen TH (2015), 'The Human Nuclear Exosome Targeting Complex Is Loaded onto Newly Synthesized RNA to Direct Early Ribonucleolysis', *Cell Rep*, 10(2): 178-192

- Lubas M, Christensen MS, Kristiansen MS, Domanski M, Falkenby LG, Lykke-Andersen S, Andersen JS, Dziembowski A, Jensen TH (2011), 'Interaction profiling identifies the human nuclear exosome targeting complex', *Mol Cell*, 43(4): 624-637

- Lubelsky Y, Sasaki T, Kuipers MA, Lucas I, Le Beau MM, Carignon S, Debatisse M, Prinz JA, Dennis JH, Gilbert DM (2011), 'Pre-replication complex proteins assemble at regions of low nucleosome occupancy within the Chinese hamster dihydrofolate reductase initiation zone', *Nucleic Acids Res*, 39(8): 3141–3155

- Luger K, Mäder AW, Richmond RK, Sargent DF, Richmond TJ (1997), 'Crystal structure of the nucleosome core particle at 2.8 A resolution', *Nature*, 389(6648): 251-260

- Luna R, Jimeno S, Marín M, Huertas P, García-Rubio M, Aguilera A (2005), 'Interdependence between transcription and mRNP processing and export, and its impact on genetic stability', *Mol Cell*, 18(6): 711-722

- Luo W, Ji Z, Pan Z, You B, Hoque M, Li W, Gunderson SI, Tian B (2013), 'The Conserved Intronic Cleavage and Polyadenylation Site of CstF-77 Gene Imparts Control of 3′ End Processing Activity through Feedback Autoregulation and by U1 snRNP', *PLoS Genet*, 9(7): e1003613

- MacAlpine HK, Gordân R, Powell SK, Hartemink AJ, MacAlpine DM (2010), 'Drosophila ORC localizes to open chromatin and marks sites of cohesin complex loading', *Genome Res*, 20(2): 201-211

- Manzo SG, Hartono SR, Sanz LA, Marinello J, De Biasi S, Cossarizza A, Capranico G, Chedin F (2018), 'DNA Topoisomerase I differentially modulates R-loops across the human genome', *Genome Biol*, 19(1): 100

- Marchese FP, Grossi E, Marín-Béjar O, Bharti SK, Raimondi I, González J, Martínez-Herrera DJ, Athie A, Amadoz A, Brosh RM Jr, Huarte M (2016), 'A Long Noncoding RNA Regulates Sister Chromatid Cohesion', *Mol Cell*, 63(3): 397-407

- Maresca TJ1, Freedman BS, Heald R (2005), 'Histone H1 is essential for mitotic chromosome architecture and segregation in Xenopus laevis egg extracts', *J Cell Biol*, 169(6): 859-869

- Martianov I, Ramadass A, Serra Barros A, Chow N, Akoulitchev A (2007), 'Repression of the human dihydrofolate reductase gene by a non-coding interfering transcript', *Nature*, 445(7128): 666-670

- Martínez-Jiménez MI, García-Gómez S, Bebenek K, Sastre-Moreno G, Calvo PA, Díaz-Talavera A, Kunkel TA, Blanco L (2015), 'Alternative solutions and new scenarios for translesion DNA synthesis by human PrimPol', *DNA Repair (Amst)*, 29: 127-138

- Maruyama A, Mimura J, Itoh K (2014), 'Non-coding RNA derived from the region adjacent to the human HO-1 E2 enhancer selectively regulates HO-1 gene induction by modulating Pol II binding', *Nucleic Acids Res*, 42(22):13599-13614

- Mechali M (2010), 'Eukaryotic DNA replication origins: many choices for appropriate answers', *Nat Rev Mol Cell Biol*, 11(10): 728-738

- Melgar MF, Collins FS, Sethupathy P (2011), 'Discovery of active enhancers through bidirectional expression of short transcripts', *Genome Biol*, 12(11): R113

- Melo CA, Drost J, Wijchers PJ, van de Werken H, de Wit E, Oude Vrielink JA, Elkon R, Melo SA, Léveillé N, Kalluri R, de Laat W, Agami R (2013), 'eRNAs Are Required for p53-Dependent Enhancer Activity and Gene Transcription', *Mol Cell*, 49(3):524-35

- Méndez J, Stillman B (2010), 'Chromatin association of human origin recognition complex, cdc6, and minichromosome maintenance proteins during the cell cycle: assembly of prereplication complexes in late mitosis', *Mol Cell Biol*, 20(22): 8602-8612

- Meola N, Domanski M, Karadoulama E, Chen Y, Gentil C, Pultz D, Vitting-Seerup K, Likke-Andersen S, Andersen JS, Sendelin A, Jensen TH (2016), 'Identification of a Nuclear Exosome Decay Pathway for Processed Transcripts', *Mol Cell*, 64(3):520-533

- Mercer TR, Gerhardt DJ, Dinger ME, Crawford J, Trapnell C, Jeddeloh JA, Mattick JS, Rinn JL (2011), 'Targeted RNA sequencing reveals the deep complexity of the human transcriptome', *Nat Biotechnol*, 30(1): 99-104

- Merkenschlager M, Nora EP (2016), 'CTCF and Cohesin in Genome Folding and Transcriptional Gene Regulation', *Annu Rev Genomics Hum Genet*, 17:17-43

- Merrikh H, Zhang Y, Grossman AD, Wang JD (2012), 'Replication–transcription conflicts in bacteria', *Nat Rev Microbiol*, 10(7): 449-458

- Meryet-Figuiere M, Alaei-Mahabadi B, Ali MM, Mitra S, Subhash S, Pandey GK, Larsson E, Kanduri C (2014), 'Temporal separation of replication and transcription during S-phase progression', *Cell Cycle*, 13(20): 3241-3248

- Mi H, Muruganujan A, Huang X, Ebert D, Mills C, Guo X, Thomas PD (2019), 'Protocol Update for large-scale genome and gene function analysis with the PANTHER classification system (v.14.0)', *Nat Protoc*, 14(3): 703-721

- Mieczkowski J, Cook A, Bowman SK, Mueller B, Alver BH, Kundu S, Deaton AM, Urban JA, Larschan E, Park PJ, Kingston RE, Tolstorukov MY (2016), 'MNase titration reveals differences between nucleosome occupancy and chromatin accessibility', *Nat Commun*, 7:11485

- Millán-Ariño L, Islam AB, Izquierdo-Bouldstridge A, Mayor R, Terme JM, Luque N, Sancho M, López-Bigas N, Jordan A (2014), 'Mapping of six somatic linker histone H1 variants in human breast cancer cells uncovers specific features of H1.2', *Nucleic Acids Res*, 42(7): 4474-4493

- Miotto B, Struhl K (2010), 'HBO1 histone acetylase activity is essential for DNA replication licensing and inhibited by geminin', *Mol Cell*, 37(1): 57-66

- Mirkin EV, Mirkin SM (2005), 'Mechanisms of Transcription-Replication Collisions in Bacteria', *Mol Cell Biol*, 25(3):888-895

- Misteli T, Gunjan A, Hock R, Bustin M, Brown DT (2000), 'Dynamic binding of histone H1 to chromatin in living cells', *Nature*, 408(6814): 877-881

- Mitchell P (2014), 'Exosome substrate targeting: the long and short of it', *Biochem Soc Trans*, 42(4): 1129-1134

- Mitchell P, Petfalski E, Shevchenko A, Mann M, Tollervey D (1997), 'The exosome: a conserved eukaryotic RNA processing complex containing multiple 3'→5' exoribonucleases', *Cell*, 91(4): 457-466

- Morales JC, Richard P, Patidar PL, Motea EA, Dang TT, Manley JL, Boothman DA (2016), 'XRN2 Links Transcription Termination to DNA Damage and Replication Stress', *PLoS Genet*, 12(7): e1006107

- Mourón S, Rodríguez-Acebes S, Martínez-Jiménez MI, García-Gómez S, Chocrón S, Blanco L, Méndez J (2013), 'Repriming of DNA synthesis at stalled replication forks by human PrimPol', *Nat Struct Mol Biol*, 20(12): 1383-1389

- Mousavi K, Zare H, Dell'Orso S, Grontved L, Gutiérrez-Cruz G, Derfoul A, Hager GL, Sartorelli V (2013), 'eRNAs Promote Transcription by Establishing Chromatin Accessibility at Defined Genomic Loci', *Mol Cell*, 51(5): 606-617

- Mueller B, Mieczkowski J, Kundu S, Wang P, Sadreyev R, Tolstorukov MY, Kingston RE (2017), 'Widespread changes in nucleosome accessibility without changes in nucleosome occupancy during a rapid transcriptional induction', *Genes Dev*, 31(5):451-462

- Mukherjee N, Calviello L, Hirsekorn A, de Pretis S, Pelizzola M, Ohler U (2017), 'Integrative classification of human coding and noncoding genes through RNA metabolism profiles', *Nat Struct Mol Biol*, 24(1): 86-96

- Murga M, Jaco I, Fan Y, Soria R, Martinez-Pastor B, Cuadrado M, Yang SM, Blasco MA, Skoultchi AI, Fernandez-Capetillo O (2007), 'Global chromatin compaction limits the strength of the DNA damage response', *J Cell Biol*, 178(7): 1101-1108

- Murray SC, Haenni S, Howe FS, Fischl H, Chocian K, Nair A, Mellor J (2015), 'Sense and antisense transcription are associated with distinct chromatin architectures across genes', *Nucleic Acids Res*, 43(16): 7823-7837

- Ntini E, Järvelin AI, Bornholdt J, Chen Y, Boyd M, Jørgensen M, Andersson R, Hoof I, Schein A, Andersen PR, Andersen PK, Preker P, Valen E, Zhao X, Pelechano V, Steinmetz LM, Sandelin A, Jensen TH (2013), 'Polyadenylation site-induced decay of upstream transcripts enforces promoter directionality', *Nat Struct Mol Biol*, 20(8): 923-928

- Ogami K, Chen Y, Manley JL (2018), 'RNA surveillance by the nuclear RNA exosome: mechanisms and significance', *Noncoding RNA*, 4(1) pii:8

- Ogami K, Richard P, Chen Y, Hoque M, Li W, Moresco JJ, Yates JR, Tian B, Manley JL (2017), 'An Mtr4/ZFC3H1 complex facilitates turnover of unstable nuclear RNAs to

prevent their cytoplasmic transport and global translational repression', *Genes Dev*, 31(12): 1257-1271

- Pan C, Fan Y (2016), 'Role of H1 linker histones in mammalian development and stem cell differentiation', *Biochim Biophys Acta*, 1859(3): 496-509

- Patro R, Duggal G, Love MI, Irizarry RA, Kingsford C (2017), 'Salmon provides fast and bias-aware quantification of transcript expression', *Nat Methods*, 14(4): 417-419

- Pefanis E, Wang J, Rothschild G, Lim J, Chao J, Rabadan R, Economides AN, Basu U (2014), 'Noncoding RNA transcription targets AID to divergently transcribed loci in B cells', *Nature*, 514(7522): 389-393

- Pefanis E, Wang J, Rothschild G, Lim J, Kazadi D, Sun J, Federation A, Chao J, Elliott O, Liu ZP, Economides AN, Bradner JE, Rabadan R, Basu U (2015), 'RNA exosome-regulated long non-coding RNA transcription controls super-enhancer activity', *Cell*, 161(4): 774-789

- Pekowska A, Benoukraf T, Zacarias-Cabeza J, Belhocine M, Koch F, Holota H, Imbert J, Andrau JC, Ferrier P, Spicuglia S (2011), 'H3K4 tri-methylation provides an epigenetic signature of active enhancers', *EMBO J*, 30(20): 4198-4210

- Pennings S, Meersseman G, Bradbury EM (1994), 'Linker histones H1 and H5 prevent the mobility of positioned nucleosomes', *Proc Natl Acad Sci USA*, 91(22): 10275-10279.

- Peterlin BM, Price DH (2006), 'Controlling the Elongation Phase of Transcription with P-TEFb', *Mol Cell*, 23(3):297-305

- Petersen BO, Lukas J, Sørensen CS, Bartek J, Helin K (1999), 'Phosphorylation of mammalian CDC6 by cyclin A/CDK2 regulates its subcellular localization', *EMBO J*, 18(2): 396–410

- Petryk N, Kahli M, d'Aubenton-Carafa Y, Jaszczyszyn Y, Shen Y, Silvain M, Thermes C, Chen CL, Hyrien O (2016), 'Replication landscape of the human genome', *Nat Commun*, 7: 10208

- Petty E, Pillus L (2013), 'Balancing chromatin remodeling and histone modifications in transcription', *Trends Genet*, 29(11): 621-629

- Picard F, Cadoret JC, Audit B, Arneodo A, Alberti A, Battail C, Duret L, Prioleau MN (2014), 'The spatiotemporal program of DNA replication is associated with specific combinations of chromatin marks in human cells', *PLoS Genet*, 10(5): e1004282

- Podhorecka M, Skladanowski A, Bozko P (2010), 'H2AX Phosphorylation: Its Role in DNA Damage Response and Cancer Therapy', *J Nucleic Acids*, pii: 920161

- Popova EY, Grigoryev SA, Fan Y, Skoultchi AI, Zhang SS, Barnstable CJ (2013), 'Developmentally regulated linker histone H1c promotes heterochromatin condensation and mediates structural integrity of rod photoreceptors in mouse retina', *J Biol Chem*, 288(24): 17895-17907

- Prado F, Aguilera A (2005), 'Impairment of replication fork progression mediates RNA polII transcription-associated recombination', *EMBO J*, 24(6): 1267–1276

- Preker P, Almvig K, Christensen MS, Valen E, Mapendano CK, Sandelin A, Jensen TH (2011), 'PROMoter uPstream Transcripts share characteristics with mRNAs and are produced upstream of all three major types of mammalian promoters', *Nucleic Acids Res*, 39(16): 7179-7193

- Prorok P, Artufel M, Aze A, Coulombe P, Peiffer I, Lacroix L, Guédin A, Mergny JL, Damaschke J, Schepers A, Ballester B, Mechali M (2019), 'Involvement of G-quadruplex regions in mammalian replication origin activity', *Nat Commun*, 10: 3274

- Proudfoot NJ (1989), 'How RNA polymerase II terminates transcription in higher eukaryotes', *Trends Biochem Sci*, 14(3): 105-110.

- Proudfoot NJ (2016), 'Transcriptional termination in mammals: Stopping the RNA polymerase II juggernaut', *Science*, 352(6291): aad9926

- Pugh BF (2010), 'A preoccupied position on nucleosomes', *Nat Struct Mol Biol*, 17(8):923

- Rabini S, Franke K, Saftig P, Bode C, Doenecke D, Drabent B (2000), 'Spermatogenesis in mice is not affected by histone H1.1 deficiency', *Exp Cell Res*, 255(1):114-124

- Ragland RL, Patel S, Rivard RS, Smith K, Peters AA, Bielinsky AK, Brown EJ (2013), 'RNF4 and PLK1 are required for replication fork collapse in ATR-deficient cells', *Genes Dev*, 27(20): 2259–2273

- Rahl PB, Lin CY, Seila AC, Flynn RA, McCuine S, Burge CB, Sharp PA, Young RA (2010), 'c-Myc regulates transcriptional pause release', *Cell*, 141(3):432-445

- Rao H, Stillman B (1995), 'The origin recognition complex interacts with a bipartite DNA binding site within yeast replicators', *Proc Natl Acad Sci USA*, 92(6): 2224–2228

- Remus D, Beuron F, Tolun G, Griffith JD, Morris EP, Diffley JFX (2009), 'Concerted loading of Mcm2-7 double hexamers around DNA during DNA replication origin licensing', 139(4): 719-730

- Ricci MA, Manzo C, García-Parajo MF, Lakadamyali M, Cosma MP (2015), 'Chromatin fibers are formed by heterogeneous groups of nucleosomes in vivo', *Cell*, 160(6): 1145-1158

- Robertson AG, Bilenky M, Tam A, Zhao Y, Zeng T, Thiessen N, Cezard T, Fejes AP, Wederell ED, Cullum R, Euskirchen G, Krzywinski M, Birol I, Snyder M, Hoodless PA, Hirst M, Marra MA, Jones SJ (2008), 'Genome-wide relationship between histone H3 lysine 4 mono- and tri-methylation and transcription factor binding', *Genome Res*, 18(12): 1906-1917

- Rondón AG, Mischo HE, Kawauchi J, Proudfoot NJ (2009), 'Fail-safe transcriptional termination for protein-coding genes in S. cerevisiae', *Mol Cell*, 36(1): 88-98

- Rutkowski AJ, Erhard F, L'Hernault A, Bonfert T, Schilhabel M, Crump C, Rosenstiel P, Efstathiou S, Zimmer R, Friedel CC, Dölken L (2015), 'Widespread disruption of host transcription termination in HSV-1 infection', *Nat Commun*, 6:7126

- Sainsbury S, Bernecky C, Cramer P (2015), 'Structural basis of transcription initiation by RNA polymerase II', *Nat Rev Mol Cell Biol*, 16(3): 129-143

- Sancho M, Diani E, Beato M, Jordan A (2008), 'Depletion of human histone H1 variants uncovers specific roles in gene expression and cell growth', *PLoS Genet*, 4(10): e1000227

- Santos-Rosa H, Schneider R, Bernstein BE, Karabetsou N, Morillon A, Weise C, Schreiber SL, Mellor J, Kouzarides T (2003), 'Methylation of histone H3 K4 mediates association of the Isw1p ATPase with chromatin', *Mol Cell*, 12(5): 1325-1332

- Sanyal A, Lajoie BR, Dekker J (2012), 'The long-range interaction landscape of gene promoters', *Nature*, 489(7414): 109-113

- Schalch T (2017), 'Higher order chromatin structures are taking shape', *Z Med Phys*, 27(2):75-77

- Schaukowitch K, Joo JY, Liu X, Watts JW, Martínez C, Kim TK (2014), 'Enhancer RNA Facilitates NELF Release from Immediate Early Genes', *Mol Cell*, 56(1): 29-42.

- Schlackow M, Nojima T, Gomes T, Dhir A, Carmo-Fonseca M, Proudfoot NJ (2017), 'Distinctive Patterns of Transcription and RNA Processing for Human lincRNAs', *Mol Cell*, 65(1): 25-38

- Schmitt S, Prestel M, Paro R (2005), 'Intergenic transcription through a polycomb group response element counteracts silencing', *Genes Dev*, 19(6): 697-708

- Schnetz MP, Handoko L, Akhtar-Zaidi B, Bartels CF, Pereira CF, Fisher AG, Adams DJ, Flicek P, Crawford GE, Laframboise T, Tesar P, Wei CL, Scacheri PC (2010), 'CHD7 targets active gene enhancer elements to modulate ES cell-specific gene expression', *PLoS Genet*, 6(7): e1001023

- Schones DE, Cui K, Cuddapah S, Roh TY, Barski A, Wang Z, Wei G, Zhao K (2008), 'Dynamic regulation of nucleosome positioning in the human genome', *Cell*, 132(5): 887-898

- Schulz D, Schwalb B, Kiesel A, Baejen C, Torkler P, Gagneur J, Soeding J, Cramer P (2013), 'Transcriptome Surveillance by Selective Termination of Noncoding RNA Synthesis', *Cell*, 155(5): 1075-1087

- Seila AC, Calabrese JM, Levine SS, Yeo GW, Rahl PB, Flynn RA, Young RA, Sharp PA (2008), 'Divergent transcription from active promoters', *Science*, 322(5909): 1849-51

- Sequeira-Mendes J, Díaz-Uriarte R, Apedaile A, Huntley D, Brockforff N, Gómez M (2009), 'Transcription Initiation Activity Sets Replication Origin Efficiency in Mammalian Cells', *PLoS Genet*, 5(4): e1000446

- Shahbazian MD, Grunstein M (2007), 'Functions of site-specific histone acetylation and deacetylation', *Annu Rev Biochem*, 76:75-100

- Shen X, Gorovsky MA (1996), 'Linker histone H1 regulates specific gene expression but not global transcription in vivo', *Cell*, 86(3): 475-483

- Sheu YJ, Stillman B (2006), 'Cdc7-Dbf4 Phosphorylates MCM Proteins via a Docking Site-Mediated Mechanism to Promote S Phase Progression', *Mol Cell*, 24(1):101-113

- Shi L, Wen H1, Shi X (2017), 'The Histone Variant H3.3 in Transcriptional Regulation and Human Disease', *J Mol Biol*, 429(13): 1934-1945

- Shivji MKK, Renaudin X, Williams ÇH, Venkitaraman AR (2018), 'BRCA2 Regulates Transcription Elongation by RNA Polymerase II to Prevent R-Loop Accumulation', *Cell Rep*, 22(4): 1031-1039

- Shlyueva D, Stampfel G, Stark A (2014), 'Transcriptional enhancers: from properties to genome-wide predictions', *Nat Rev Genet*, 15(4): 272-286

- Sims RJ 3rd, Millhouse S, Chen CF, Lewis BA, Erdjument-Bromage H, Tempst P, Manley JL, Reinberg D (2008), 'Recognition of trimethylated histone H3 lysine 4 facilitates the recruitment of transcription postinitiation factors and pre-mRNA splicing', *Mol Cell*, 28(4): 665-676

- Sirotkin AM, Edelmann W, Cheng G, Klein-Szanto A, Kucherlapati R, Skoultchi AI (1995), 'Mice develop normally without the H1(0) linker histone', *Proc Natl Acad Sci USA*, 92(14): 6434-6438

- Skourti-Stathaki K, Proudfoot NJ (2014), 'A double-edged sword: R loops as threats to genome integrity and powerful regulators of gene expression', *Genes Dev*, 28(13): 1384-1396

- Smirnov E, Borkovec J, Kováčik L, Svidenská S, Schröfel A, Skalníková M, Švindrych Z, Křížek P, Ovesný M, Hagen GM, Juda P, Michalová K, Cardoso C, Cmarko D, Raška I (2014), 'Separation of replication and transcription domains in nucleoli', *J Struct Biol*, 188(3): 259-266

- Smolle M, Workman JL, Venkatesh S (2013), 'reSETting chromatin during transcription elongation', *Epigenetics*, 8(1): 10-15

- Sobhy H, Kumar R, Lewerentz J, Lizana L, Stenberg P (2019), 'Highly interacting regions of the human genome are enriched with enhancers and bound by DNA repair proteins', *Sci Rep*, 9(1):4577

- Stirling PC, Chan YA, Minaker SW, Aristizabal MJ, Barrett I, Sipahimalani P, Kobor MS, Hieter P (2012), 'R-loop-mediated genome instability in mRNA cleavage and polyadenylation mutants', *Genes Dev*, 26(2): 163-175

- Sugimoto N, Tatsumi Y, Tsurumi T, Matsukage A, Kiyono T, Nishitani H, Fujita M (2004), 'Cdt1 phosphorylation by cyclin A-dependent kinases negatively regulates its function without affecting geminin binding', *J Biol Chem*, 279(19):19691-19697

- Sun M, Gadad SS, Kim DS, Kraus WL (2015), 'Discovery, Annotation, and Functional Analysis of Long Noncoding RNAs Controlling Cell-Cycle Gene Expression and Proliferation in Breast Cancer Cells', *Mol Cell*, 59(4): 698-711

- Šviković S, Crisp A, Tan-Wong SM, Guilliam TA, Doherty AJ, Proudfoot NJ, Guilbaud G, Sale JE (2019), 'R-loop formation during S phase is restricted by PrimPol-mediated repriming', *EMBO J*, 38(3)

- Szerlong HJ, Herman JA, Krause CM, DeLuca JG, Skoultchi A, Winger QA, Prenni JE, Hansen JC (2015), 'Proteomic characterization of the nucleolar linker histone H1 interaction network', *J Mol Biol*, 427(11): 2056-2071

- Takemata N, Oda A, Yamada T, Galipon J, Miyoshi T, Suzuki Y, Sugano S, Hoffman CS, Hirota K, Ohta K (2016), 'Local potentiation of stress-responsive genes by upstream noncoding transcription', *Nucleic Acids Res*, 44(11): 5174-5189

- Takeuchi A, Iida K, Tsubota T, Hosokawa M, Denawa M, Brown JB, Ninomiya K, Ito M, Kimura H, Abe T, Kiyonari H, Ohno K, Hagiwara M (2018), 'Loss of Sfpq Causes Long-Gene Transcriptopathy in the Brain', *Cell Rep*, 23(5): 1326-1341

- Talbert PB, Henikoff S (2017), 'Histone variants on the move: substrates for chromatin dynamics', *Nat Rev Mol Cell Biol*, 18(2): 115-126

- Tang Z, Chen WY, Shimada M, Nguyen UT, Kim J, Sun XJ, Sengoku T, McGinty RK, Fernandez JP, Muir TW, Roeder RG (2013), 'SET1 and p300 act synergistically, through coupled histone modifications, in transcriptional activation by p53', *Cell*, 154(2): 297-310

- Tardat M, Brustel J, Kirsh O, Lefevbre C, Callanan M, Sardet C, Julien E (2010), 'The histone H4 Lys 20 methyltransferase PR-Set7 regulates replication origins in mammalian cells', *Nat Cell Biol*, 12(11): 1086-1093

- Teloni F, Michelena J, Lezaja A, Kilic S, Ambrosi C, Menon S, Dobrovolna J, Imhof R, Janscak P, Baubec T, Altmeyer M (2019), 'Efficient Pre-mRNA Cleavage Prevents Replication-Stress-Associated Genome Instability', *Mol Cell*, 73(4): 670-683

- Tennyson CN, Klamut HJ, Worton RG (1995), 'The human dystrophin gene requires 16 hours to be transcribed and is cotranscriptionally spliced', *Nat Genet*, ;9(2): 184-190

- Thoma F, Koller T, Klug A (1979), 'Involvement of histone H1 in the organization of the nucleosome and of the salt-dependent superstructures of chromatin', *J Cell Biol*, 83(2 Pt 1): 403-427

- Thorslund T, Ripplinger A, Hoffmann S, Wild T, Uckelmann M, Villumsen B, Narita T, Sixma TK, Choudhary C, Bekker-Jensen S, Mailand N (2015), 'Histone H1 couples initiation and amplification of ubiquitin signalling after DNA damage', *Nature*, 527(7578): 389-393

- Th'ng JP, Guo XW, Swank RA, Crissman HA, Bradbury EM (1994), 'Inhibition of histone phosphorylation by staurosporine leads to chromosome decondensation', *J Biol Chem*, 269(13): 9568-9573

- Tous C, Aguilera A (2007), 'Impairment of transcription elongation by R-loops in vitro', *Biochem Biophys Res Commun*, 360(2): 428-432

- Touchon M, Nicolay S, Audit B, Brodie fo Brodie EB, d'Aubenton-Carafa Y, Arneodo A, Thermes C (2005), *Proc Natl Acad Sci USA*, 102(28): 9836–9841

- Trapnell C, Williams BA, Pertea G, Mortazavi A, Kwan G, van Baren MJ, Salzberg SL, Wold BJ, Pachter L (2010), 'Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation', *Nat Biotechnol*, 28(5): 511-515

- Tudek A, Schmid M, Makaras M, Barrass JD, Beggs JD, Jensen TH (2018), 'A Nuclear Export Block Triggers the Decay of Newly Synthesized Polyadenylated RNA', *Cell Rep*, 24(9): 2457-2467

- Tuduri S, Crabbé L, Conti C, Tourrière H, Holtgreve-Grez H, Jauch A, Pantesco V, De Vos J, Thomas A, Theillet C, Pommier Y, Tazi J, Coquelle A, Pasero P (2009), 'Topoisomerase I suppresses genomic instability by preventing interference between replication and transcription', *Nat Cell Biol*, 11(11): 1315-1324

- Valton AL, Hassan-Zadeh V, Lema I, Boggetto N, Alberti P, Saintomé C, Riou JF, Prioleau MN (2014), 'G4 motifs affect origin positioning and efficiency in two vertebrate replicators', *EMBO J*, 33(7): 732–746

- Venkatesh S, Workman JL (2015), 'Histone exchange, chromatin structure and the regulation of transcription', *Nat Rev Mol Cell Biol*, 16(3): 178-189

- Vilborg A, Sabath N, Wiesel Y, Nathans J, Levy-Adam F, Yario TA, Steitz JA, Shalgi R (2017), 'Comparative analysis reveals genomic features of stress-induced transcriptional readthrough', *Proc Natl Acad Sci USA*, 114(40): E8362-E8371

- Visel A, Blow MJ, Li Z, Zhang T, Akiyama JA, Holt A, Plajzer-Frick I, Shoukry M, Wright C, Chen F, Afzal V, Ren B, Rubin EM, Pennacchio LA (2009), 'ChIP-seq accurately predicts tissue-specific activity of enhancers', *Nature*, 457(7231): 854-858

- Vo N, Goodman RH (2001), 'CREB-binding protein and p300 in transcriptional regulation', *J Biol Chem*, 276(17): 13505-13508

- Vujatovic O, Zaragoza K, Vaquero A, Reina O, Bernués J, Azorín F (2012), 'Drosophila melanogaster linker histone dH1 is required for transposon silencing and to preserve genome integrity', *Nucleic Acids Res*, 40(12): 5402-5414

- Wahba L, Amon JD, Koshland D, Vuica-Ross M (2011), 'RNase H and multiple RNA biogenesis factors cooperate to prevent RNA:DNA hybrids from generating genome instability', *Mol Cell*, 44(6): 978-988

- Wang KC, Yang YW, Liu B, Sanyal A, Corces-Zimmerman R, Chen Y, Lajoie BR, Protacio A, Flynn RA, Gupta RA, Wysocka J, Lei M, Dekker J, Helms JA, Chang HY (2011), 'A long noncoding RNA maintains active chromatin to coordinate homeotic gene expression', *Nature*, 472(7341):120-124

- Wang R, Wang J, Paul AM, Acharya D, Bai F, Huang F, Guo YL (2013), 'Mouse embryonic stem cells are deficient in type I interferon expression in response to viral infections and double-stranded RNA', *J Biol Chem*, 288(22): 15926-15936

- Wang R, Zheng D, Wei L, Ding Q, Tian B (2019), 'Regulation of Intronic Polyadenylation by PCF11 Impacts mRNA Expression of Long Genes', *Cell Rep*, 26(10): 2766-2778

- Wang SW, Stevenson AL, Kearsey SE, Watt S, Bähler J (2008), 'Global role for polyadenylation-assisted nuclear RNA degradation in posttranscriptional gene silencing', *Mol Cell Biol*, 28(2): 656-665

- Wang Z, Zang C, Rosenfeld JA, Schones DE, Barski A, Cuddapah S, Cui K, Roh TY, Peng W, Zhang MQ, Zhao K (2008), 'Combinatorial patterns of histone acetylations and methylations in the human genome', *Nat Genet*, 40(7): 897-903

- Webb S, Hector RD, Kudla G, Granneman S (2014), 'PAR-CLIP data indicate that Nrd1-Nab3-dependent transcription termination regulates expression of hundreds of protein coding genes in yeast', *Genome Biol*, 15(1):R8

- Weber CM, Ramachandran S, Henikoff S (2014), 'Nucleosomes Are Context-Specific, H2A.Z-Modulated Barriers to RNA Polymerase', *Mol Cell*, 53(5): 819-830

- Werner MS, Ruthenburg AJ (2015), 'Nuclear Fractionation Reveals Thousands of Chromatin-Tethered Noncoding RNAs Adjacent to Active Genes', *Cell Rep*, 12(7): 1089-1098

- Werner MS, Sullivan MA, Shah RN, Nadadur RD, Grzybowski AT, Galat V, Moskowitz IP, Ruthenburg A (2017), 'Chromatin-enriched lncRNAs can act as cell-type specific activators of proximal gene transcription', *Nat Struct Mol Biol*, 24(7): 596-603

- West JA, Cook A, Alver BH, Stadtfeld M, Deaton AM, Hochedlinger K, Park PJ, Tolstorukov MY, Kingston RE (2014), 'Nucleosomal occupancy changes locally over key regulatory regions during cell differentiation and reprogramming', *Nat Commun*, 5: 4719

- Wight M, Werner A (2013), 'The functions of natural antisense transcripts', *Essays Biochem*, 54: 91-101

- Witteveldt J, Knol LI, Macias S (2019), 'MicroRNA-deficient mouse embryonic stem cells acquire a functional interferon response', *Elife*, 8. pii: e44171

- Woodcock CL, Skoultchi AI, Fan Y (2006), 'Role of linker histone in chromatin structure and function: H1 stoichiometry and nucleosome repeat length', *Chromosome Res*, 14(1): 17-25

- Wysocka J, Swigut T, Milne TA, Dou Y, Zhang X, Burlingame AL, Roeder RG, Brivanlou AH, Allis CD (2005), 'WDR5 associates with histone H3 methylated at K4 and is essential for H3 K4 methylation and vertebrate development', *Cell*, 121(6): 859-872

- Yang SM, Kim BJ, Norwood Toro L, Skoultchi AI (2013), 'H1 linker histone promotes epigenetic silencing by regulating both DNA methylation and histone H3 methylation', *Proc Natl Acad Sci USA*, 110(5): 1708-1713

- Yeeles JTP, Poli J, Marians KJ, Pasero P (2013), 'Rescuing Stalled or Damaged Replication Forks', *Cold Spring Harb Perspect Biol*, 5(5): a012815

- Yu X, Li Z (2015), 'Long non-coding RNA HOTAIR: A novel oncogene', *Mol Med Rep*, 12(4): 5611-5618

- Zabidi MA, Arnold CD, Schernhuber K, Pagani M, Rath M, Frank O, Stark A (2015), 'Enhancer-core-promoter specificity separates developmental and housekeeping gene regulation', *Nature*, 518(7540): 556-559

- Zeitlinger J, Stark A, Kellis M, Hong JW, Nechaev S, Adelman K, Levine M, Young RA (2007), 'RNA polymerase stalling at developmental control genes in the Drosophila melanogaster embryo', *Nat Genet*, 39(12): 1512-1516

- Zhang H, Rigo F, Martinson HG (2015), 'Poly(A) Signal-Dependent Transcription Termination Occurs through a Conformational Change Mechanism that Does Not Require Cleavage at the Poly(A) Site', *Mol Cell*, 59(3): 437-448

- Zhang P, Branson OE, Freitas MA, Parthun MR (2016), 'Identification of replication-dependent and replication-independent linker histone complexes: Tpr specifically promotes replication-dependent linker histone stability', *BMC Biochem*, 17(1):18

-Zhang Y, Cooke M, Panjwani S, Cao K, Krauth B, Ho PY, Medrzycki M, Berhe DT, Pan C, McDevitt TC, Fan Y (2012), 'Histone h1 depletion impairs embryonic stem cell differentiation', *PLoS Genet*, 8(5): e1002691

- Zhang Y, He Q, Hu Z, Feng Y, Fan L, Tang Z, Yuan J, Shan W, Li C, Hu X, Tanyi JL, Fan Y, Huang Q, Montone K, Dang CV, Zhang L (2016), 'Long noncoding RNA LINP1 regulates repair of DNA double-strand breaks in triple-negative breast cancer', Nat *Struct Mol Biol*, 23(6): 522-530

- Zhao Y, Garcia BA (2015), 'Comprehensive Catalog of Currently Documented Histone Modifications', *Cold Spring Harb Perspect Biol*, 7(9): a025064

- Zheng Y, John S, Pesavento JJ, Schultz-Norton JR, Schiltz RL, Baek S, Nardulli AM, Hager GL, Kelleher NL, Mizzen CA (2010), 'Histone H1 phosphorylation is associated with transcription by RNA polymerases I and II', *J Cell Biol*, 189(3): 407-415

- Zhu Y, Sun L, Chen Z, Whitaker JW, Wang T, Wang W (2013), 'Predicting enhancer transcription and activity from chromatin modifications', *Nucleic Acids Res*, 41(22): 10032-10043

- Zinder JC, Lima CD (2017), 'Targeting RNA for processing or destruction by the eukaryotic RNA exosome and its cofactors', *Genes Dev*, 31(2): 88-100

- Zou L, Stillman B (1998), 'Formation of a preinitiation complex by S-phase cyclin CDK-dependent loading of Cdc45p onto chromatin', *Science*, 280(5363): 593-596

ANNEX I

Supplementary tables

Supplementary Table S1: Number of total and aligned reads and replication origin peaks in the SNS-seq experiments.

	Total reads	*Aligned reads (mm10)*	*ORI number*	*Number of common ORIs between replicates*
WT I	140536654	73122129	94590	73792
WT II	247367362	164720257	106891	
H1-TKO I	155418772	121941851	56615	33215
H1-TKO II	191209482	141877552	91273	

Supplementary Table S2: Number of total and aligned reads in cheRNA-seq experiments. The alignment was performed against the mouse genome and the luciferase spike-in coding sequence.

	Total reads	*Aligned reads (mm10)*	*Aligned reads (luciferase sequence)*	*Ratio luciferase/aligned* ($\cdot 10^6$)
WT I	69987640	60689147	29090	479.3
WT II	75114669	67209489	57199	851.1
WT III	64235353	56881476	54865	964.5
H1-TKO I	72937563	62737821	69180	1102.7
H1-TKO II	72006537	63748022	50290	788.9
H1-TKO III	73957462	65828537	68994	1048.1

Supplementary Table S3: Number of total and aligned reads in RNApolII ChIP-seq experiments. The alignment was performed against the mouse mm10 genome and the human hg19 genome (see Materials and Methods)

	Total reads	*Aligned mouse reads (mm10)*	*Aligned human reads (hg19)*	*Ratio human/aligned*
WT	76467921	62366120	5313378	7.9%
H1-TKO	64202877	51396457	3901802	7.0%

Supplementary Table S4: Number of total and aligned reads in MNase-seq experiments. A fragment was considered correct when both reads aligned to the same chromosome and were separated by less than 250 bp (see Materials and Methods)

	Total paired-end reads	*Aligned mouse fragments (mm10)*
WT I	185253174 (x2)	172722043
WT II	144299141 (x2)	138043747
H1-TKO I	170218301 (x2)	160109932
H1-TKO II	179383961 (x2)	172246890

Supplementary Table S5: Numerical values and statistic parameters of fork rate in shSCR and shExosc3 cells, corresponding to the scatter plots in Figure 26.

	shSCR			***shExosc3***		
	Untreated	*DRB*	*DRB release*	*Untreated*	*DRB*	*DRB release*
Number of values	200	243	159	148	161	169
Minimum	0.3322	0.3284	0.2576	0.3042	0.288	0.278
25% percentile	0.7358	0.6839	0.7954	0.5328	0.7228	0.6075
Median	0.8941	0.9349	0.9195	0.6578	0.9759	0.7569
75% percentile	1.074	1.2	1.195	0.8043	1.155	0.9367
Maximum	1.968	1.833	1.945	1.272	1.96	1.435
Mean	0.94	0.9471	0.9823	0.6837	0.9743	0.7755
Std deviation	0.2981	0.3151	0.3309	0.2003	0.337	0.234
Std error	0.02108	0.02022	0.02625	0.01647	0.02656	0.018
p-value	-	0.6905	0.2803	<0.0001	<0.0001	<0.0001

Supplementary Table S6: List of qPCR primer pairs

Primer name	*Sequence (5' > 3')*	*Anneal Temperature*	*Amplicon size*
Kcnq1ot1-Fw	TTGGATTACTTCGGTGGGCT	60ºC	129 bp
Kcnq1ot1-Rv	ACACGGATGAAAACCACGCT		
NEAT1-Fw	TTGGGACAGTGGACGTGTGG	60ºC	206 bp
NEAT1-Rv	TCAAGTGCCAGCAGACAGCA		
Klf16-Fw	GTGTACCAAGCGGTTCACC	60ºC	88 bp
Klf16-Rv	CAGGTCGTCGCAGGAGTTC		
Nat8L-Fw	TGTGCATCCGCGAGTTCCGC	60ºC	94 bp
Nat8L-Rv	GCGGAAAGCCGTGTTGGGGA		
Luciferase-Fw	ACACCCGAGGGGGATGATAA	60ºC	75 bp
Luciferase-Rv	CCAGATCCACAACCTTCGCT		
Haus 105	CTGGCTCACAGTACCTTCAG	60ºC	63 bp
Haus 106	GTGATGACTGCGTGGTCAAG		
Haus 110	GGTGAGCTTTTAGCAGTGTG	60ºC	64 bp
Haus 111	CTGAACTACAACTGGTTCACG		
VpsC3_9	TTCTCGGTAGGGAGTGGAAG	60ºC	68 bp
VpsC3_10	TACTCAGGACCAGAAGCCAG		
Vps115	TTCCGGTCGCCAAAGCCTC	60ºC	52 bp
Vps116	CCCAGTATCGGAGCTACCCG		
Taq-Ex-Fw	GACTTCCGCCACCCCTG	62ºC	149 bp
Taq-Ex-Rv	AGCTAACGGAGTTTCGGCTT		
Taq-In-Fw	GCCTGGAGCTTGAGAGAGTC	60ºC	64 bp
Taq-In-Rv	GGATAACGAAAACCGAAGGA		
Lma-Ex-Fw	TGAGTGGTGACGCTTGTAAG	58ºC	142 bp
Lma-Ex-Rv	AGATCTACTGATGCGACTGC		
Lma-In-Fw	GGCAAAGTGAACAGCATGTAG	60 ºC	64 bp
Lma-In-Rv	CTGGAGAGAATACAATGCACTG		

ANNEX II

Publication list

- Lombraña R, Álvarez A, **Fernández-Justel JM**, Almeida R, Poza-Carrión C, Gomes F, Calzada A, Requena JM and Gómez M (2016), 'Transcriptionally driven DNA replication programme of the human parasite *Leishmania major*', *Cell Rep* 16(6): 1774-1786

- Almeida R[1], **Fernández-Justel JM[1]**, Santa-María C[1], Cadoret JC, Cano-Aroca L, Lombraña R, Herranz G, Agresti A, Gómez M (2018), 'Chromatin conformation regulates the coordination between DNA replication and transcription', *Nat Commun*, 9(1): 1590
  [1]Co-first authors

- Massip F, Laurent M, Brossas C, **Fernández-Justel JM**, Gómez M, Prioleau MN, Duret L, Picard F (2019), 'Evolution of replication origins in vertebrate genomes: rapid turnover despite selective constraints', *Nucleic Acids Res*, 47(10): 5114-5125